Entropy and the fourth moment phenomenon

Ivan Nourdin, Giovanni Peccati and Yvik Swan

April 5, 2013

Abstract

We develop a new method for bounding the relative entropy of a random vector in terms of its Stein factors. Our approach is based on a novel representation for the score function of smoothly perturbed random variables, as well as on the de Bruijn's identity of information theory. When applied to sequences of functionals of a general Gaussian field, our results can be combined with the Carbery-Wright inequality in order to yield multidimensional entropic rates of convergence that coincide, up to a logarithmic factor, with those achievable in smooth distances (such as the 1-Wasserstein distance). In particular, our findings settle the open problem of proving a quantitative version of the multidimensional *fourth moment theorem* for random vectors having chaotic components, with explicit rates of convergence in total variation that are independent of the order of the associated Wiener chaoses. The results proved in the present paper are outside the scope of other existing techniques, such as for instance the multidimensional Stein's method for normal approximations.

Keywords: Carbery-Wright Inequality; Central Limit Theorem; De Bruijn's Formula; Fisher Information; Fourth Moment Theorem; Gaussian Fields; Relative Entropy; Stein factors.
MSC 2010: 94A17; 60F05; 60G15; 60H07

Contents

1 Introduction

1.1 Overview and motivation

The aim of this paper is to develop a new method for controlling the relative entropy of a general random vector with values in $\mathbb{R}^d$, and then to apply this technique to settle a number of open questions concerning central limit theorems (CLTs) on a Gaussian space. Our approach is based on a fine analysis of the (multidimensional) *de Bruijn's formula*, which provides a neat representation of the derivative of the relative entropy (along the Ornstein-Uhlenbeck semigroup) in terms of the Fisher information of some perturbed random vector – see e.g. [1, 4, 5, 19, 21]. The main tool developed in this paper (see Theorem 2.10 as well as relation (1.14)) is a new powerful representation of relative entropies in terms of *Stein factors*. Roughly speaking, Stein factors are random variables verifying a generalised integration by parts formula (see (2.43) below): these objects naturally appear in the context of the multidimensional *Stein's method* for normal approximations (see e.g. [15, 33]), and implicitly play a crucial role in many probabilistic limit theorems on Gaussian or other spaces (see e.g. [33, Chapter 6], as well as [31, 34, 37]).

The study of the classical CLT for sums of independent random elements by entropic methods dates back to Linnik's seminal paper [28]. Among the many fundamental contributions to this line of research, we cite [2, 3, 5, 8, 20, 9, 10, 12, 21] (see the monograph [19] for more details on the history of the theory). All these influential works revolve around a deep analysis of the effect of analytic convolution on the creation of entropy: in this respect, a particularly powerful tool are the 'entropy jump inequalities' proved and exploited e.g. in [3, 5, 20, 6]. As discussed e.g. in [6], entropy jump inequalities are directly connected with challenging open questions in convex geometry, like for instance the *Hyperplane* and *KLS* conjectures. One of the common traits of all the above references is that they develop tools to control the Fisher information and use the aforementioned de Bruijn's formula to translate the bounds so obtained into bounds on the relative entropy.

One of the main motivations of the present paper is to initiate a systematic information-theoretical analysis of a large class of CLTs that has emerged in recent years in connection with different branches of modern stochastic analysis. These limit theorems typically involve: (a) an underlying infinite dimensional Gaussian field $\mathbf{G}$ (like for instance a Wiener process), (b) a sequence of rescaled centered random vectors $F_n = F_n(\mathbf{G})$, $n \geqslant 1$, having the form of some highly non-linear functional of the field $\mathbf{G}$. For example, each F_n may be defined as some collection of polynomial transformations of $\mathbf{G}$, possibly depending on a parameter that is integrated with respect to a deterministic measure (but much more general forms are possible). Objects of this type naturally appear e.g. in the high-frequency analysis of random fields on homogeneous spaces [29], fractional processes [30, 39], Gaussian polymers [49], or random matrices [13, 32].

In view of their intricate structure, it is in general not possible to meaningfully represent the vectors F_n in terms of some linear transformation of independent (or weakly dependent) vectors, so that the usual analytical techniques based on stochastic independence and convolution (or mixing) cannot be applied. To overcome these difficulties, a recently developed line of research (see [33] for an introduction) has revealed that, by using tools from infinite-dimensional Gaussian analysis (e.g. the so-called *Malliavin calculus*

of variations – see [38]) and under some regularity assumptions on F_n, one can control the distance between the distribution of F_n and that of some Gaussian target by means of quantities that are no more complex than the fourth moment of F_n. The regularity assumptions on F_n are usually expressed in terms of the projections of each F_n on the eigenspaces of the Ornstein-Uhlenbeck semigroup associated with $\mathbf{G}$ [33, 38].

In this area, the most prominent contribution is arguably the following one-dimensional inequality established by the first two authors (see, e.g., [33, Theorem 5.2.6]): let f be the density of a random variable F, assume that $\int_{\mathbb{R}} x^2 f(x)dx = 1$, and that F belongs to the qth eigenspace of the Ornstein-Uhlenbeck semigroup of $\mathbf{G}$ (customarily called the qth *Wiener chaos* of $\mathbf{G}$), then

$$\frac{1}{4}\int_{\mathbb{R}} |f(x) - \phi_1(x)|\, dx \leqslant \sqrt{\frac{1}{3} - \frac{1}{3q}} \times \sqrt{\int_{\mathbb{R}} x^4(f(x) - \phi_1(x))dx}, \tag{1.1}$$

where $\phi_1(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard Gaussian density (one can prove that $\int_{\mathbb{R}} x^4(f(x) - \phi_1(x))dx > 0$ for f as above). Note that $\int_{\mathbb{R}} x^4\phi_1(x)dx = 3$. A standard use of hypercontractivity therefore allows one to deduce the so-called *fourth moment theorem* established in [40]: *for a rescaled sequence* $\{F_n\}$ *of random variables living inside the qth Wiener chaos of* $\mathbf{G}$*, one has that* F_n *converges in distribution to a standard Gaussian random variable if and only if the fourth moment of* F_n *converges to* 3 *(and in this case the convergence is in the sense of total variation).* See also [39].

The quantity $\sqrt{\int_{\mathbb{R}} x^4(f(x) - \phi_1(x))dx}$ appearing in (1.1) is often called the *kurtosis* of the density f: it provides a rough measure of the discrepancy between the 'fatness' of the tails of f and ϕ_1. The systematic emergence of the normal distribution from the reduction of kurtosis, in such a general collection of probabilistic models, is a new phenomenon that we barely begin to understand. A detailed discussion of these results can be found in [33, Chapters 5 and 6]. M. Ledoux [25] has recently proved a striking extension of the fourth moment theorem to random variables living in the eigenspaces associated with a general Markov operator, whereas references [17, 22] contain similar statements in the framework of free probability.

The estimate (1.1) is obtained by combining the Malliavin calculus of variations with the Stein's method for normal approximations [15, 33]. Stein's method can be roughly described as a collection of analytical techniques, allowing one to measure the distance between random elements by controlling the regularity of the solutions to some specific ordinary (in dimension 1) or partial (in higher dimensions) differential equations. The needed estimates are often expressed in terms of the same Stein factors that lie at the core of the present paper (see Section 2.3 for definitions). It is important to notice that the strength of these techniques significantly breaks down when dealing with normal approximations in dimension *strictly* greater than 1. For instance, in view of the structure of the associated PDEs, for the time being there is no way to directly use Stein's method in order to obtain bounds in the multidimensional total variation distance (see [14, 43]). In contrast, the results of this paper allow one to deduce a number of information-theoretical generalisations of (1.1) that are valid in any dimension. It is somehow remarkable that our techniques make a pervasive use of Stein factors, without ever applying Stein's method. As an illustration, we present here a multidimensional entropic fourth moment bound that will be proved in full generality in Section 4. For $d \geqslant 1$, we write $\phi_d(\mathbf{x}) = \phi_d(x_1, ..., x_d)$ to indicate the Gaussian density $(2\pi)^{-d/2}\exp(-(x_1^2 + \cdots + x_d^2)/2)$, $(x_1, ..., x_d) \in \mathbb{R}^d$. From now on, every random object is assumed to be defined on a common probability space $(\Omega, \mathcal{F}, P)$, with E denoting expectation with respect to P.

Theorem 1.1 (Entropic fourth moment bound) *Let* $F_n = (F_{1,n}, ..., F_{d,n})$ *be a sequence of d-dimensional random vectors such that:* (i) $F_{i,n}$ *belongs to the* q_i*th Wiener*

chaos of $\mathbf{G}$*, with* $1 \leqslant q_1 \leqslant q_2 \leqslant \cdots \leqslant q_d$*;* (ii) *each* $F_{i,n}$ *has variance* 1, (iii) $E[F_{i,n}F_{j,n}] = 0$ *for* $i \neq j$*, and* (iv) *the law of* F_n *admits a density* f_n *on* $\mathbb{R}^d$*. Write*

$$\Delta_n := \int_{\mathbb{R}^d} \|\mathbf{x}\|^4 (f_n(\mathbf{x}) - \phi_d(\mathbf{x})) d\mathbf{x},$$

where $\|\cdot\|$ *stands for the Euclidean norm, and assume that* $\Delta_n \to 0$*, as* $n \to \infty$*. Then,*

$$\int_{\mathbb{R}^d} f_n(\mathbf{x}) \log \frac{f_n(\mathbf{x})}{\phi_d(\mathbf{x})} d\mathbf{x} = O(1)\,\Delta_n |\log \Delta_n|, \tag{1.2}$$

where $O(1)$ *stands for a bounded numerical sequence, depending on* $d, q_1, ..., q_d$ *and on the sequence* $\{F_n\}$*.*

As in the one-dimensional case, one has always that $\Delta_n > 0$ for f_n as in the previous statement. The quantity of the left-hand-side of (1.2) equals of course the *relative entropy* of f_n. In view of the Csiszar-Kullback-Pinsker inequality (see [16, 23, 42]), according to which

$$\int_{\mathbb{R}^d} f_n(\mathbf{x}) \log \frac{f_n(\mathbf{x})}{\phi_d(\mathbf{x})} d\mathbf{x} \geqslant \frac{1}{2} \left(\int_{\mathbb{R}^d} \Big| f_n(\mathbf{x}) - \phi_d(\mathbf{x}) \Big| d\mathbf{x} \right)^2, \tag{1.3}$$

relation (1.2) then translates in a bound on the square of the total variation distance between f_n and ϕ_d, where the dependence in Δ_n hinges on the order of the chaoses only via a multiplicative constant. This bound agrees up to a logarithmic factor with the estimates in smoother distances established in [37] (see also [35]), where it is proved that there exists a constant $K_0 = K_0(d, q_1, ..., q_d)$ such that

$$\mathbf{W}_1(f_n, \phi_d) \leqslant K_0\, \Delta_n^{1/2},$$

where $\mathbf{W}_1$ stands for the usual Wasserstein distance of order 1. Relation (1.2) also drastically improves the bounds that can be deduced from [31], yielding that, as $n \to \infty$,

$$\int_{\mathbb{R}^d} \Big| f_n(\mathbf{x}) - \phi_d(\mathbf{x}) \Big| d\mathbf{x} = O(1)\, \Delta_n^{\alpha_d},$$

where α_d is any strictly positive number verifying $\alpha_d < \frac{1}{1+(d+1)(3+4d(q_d-1))}$, and the symbol $O(1)$ stands again for some bounded numerical sequence. The estimate (1.2) seems to be largely outside the scope of any other available technique. Our results will also show that convergence in relative entropy is a necessary and sufficient condition for CLTs involving random vectors whose components live in a fixed Wiener chaos. As in [31, 36], an important tool for establishing our main results is the Carbery-Wright inequality [11], providing estimates on the small ball probabilities associated with polynomial transformations of Gaussian vectors. Observe also that, via the Talagrand's transport inequality [48], our bounds trivially provide estimates on the 2-Wasserstein distance $\mathbf{W}_2(f_n, \phi_d)$ between f_n and ϕ_d, for every $d \geqslant 1$.

We stress that, albeit our principal motivation comes from asymptotic problems on a Gaussian space, the methods developed in Section 2 are general. In fact, at the heart of the present work lie the powerful equivalences (2.45)– (2.46) (which can be considered as a new form of so-called *Stein identities*) that are valid under very weak assumptions on the target density; it is also easy to uncover a wide variety of extensions and generalizations so that we expect that our tools can be adapted to deal with a much wider class of multidimensional distributions.

The connection between Stein identities and information theory has already been noted in the literature (although only in dimension 1). For instance, explicit applications

are known in the context of Poisson and compound Poisson approximations [7, 45], and recently several promising identities have been discovered for some discrete [26, 44] as well as continuous distributions [24, 27, 41]. However, with the exception of [27], the existing literature seems to be silent about any connection between entropic CLTs and Stein's identities for normal approximations. To the best of our knowledge, together with [27] (which however focusses on bounds of a completely different nature) the present paper contains the first relevant study of the relations between the two topics.

Remark on notation. Given random vectors X, Y with values in $\mathbb{R}^d$ ($d \geqslant 1$) and densities f_X, f_Y, respectively, we shall denote by $\mathbf{TV}(f_X, f_Y)$ and $\mathbf{W}_1(f_X, f_Y)$ the *total variation* and 1-*Wasserstein* distances between f_X and f_Y (and thus between the laws of X and Y). Recall that we have the representations

$$\begin{aligned}\mathbf{TV}(f_X, f_Y) &= \sup_{A \in \mathscr{B}(\mathbb{R}^d)} \Big| P[X \in A] - P[Y \in A] \Big| \\ &= \frac{1}{2} \sup_{\|h\|_\infty \leqslant 1} \Big| E[h(X)] - E[h(Y)] \Big| \\ &= \frac{1}{2} \int_{\mathbb{R}^d} \Big| f_X(\mathbf{x}) - f_Y(\mathbf{x}) \Big| d\mathbf{x} =: \frac{1}{2} \|f_X - f_Y\|_1,\end{aligned} \tag{1.4}$$

where (here and throughout the paper) $d\mathbf{x}$ is shorthand for the Lebesgue measure on $\mathbb{R}^d$, as well as

$$\mathbf{W}_1(f_X, f_Y) = \sup_{h \in \mathrm{Lip}(1)} \Big| E[h(X)] - E[h(Y)] \Big|.$$

In order to simplify the discussion, we shall sometimes use the shorthand notation

$$\mathbf{TV}(X, Y) = \mathbf{TV}(f_X, f_Y) \text{ and } \mathbf{W}_1(X, Y) = \mathbf{W}_1(f_X, f_Y).$$

It is a well-known fact the the topologies induced by $\mathbf{TV}$ and $\mathbf{W}_1$, over the class of probability measures on $\mathbb{R}^d$, are strictly stronger than the topology of convergence in distribution (see e.g. [18, Chapter 11] or [33, Appendix C]). Finally, we agree that every logarithm in the paper has base e.

To enhance the readability of the text, the next Subsection 1.2 contains an intuitive description of our method in dimension one.

1.2 Illustration of the method in dimension one

Let F be a random variable with density $f : \mathbb{R} \to [0, \infty)$, and let Z be a standard Gaussian random variable with density ϕ_1. We shall assume that $E[F] = 0$ and $E[F^2] = 1$, and that Z and F are stochastically independent. As anticipated, we are interested in bounding the relative entropy of F (with respect to Z), which is given by the quantity

$$D(F\|Z) = \int_{\mathbb{R}} f(x) \log(f(x)/\phi_1(x)) dx.$$

Recall also that, in view of the Pinsker-Csiszar-Kullback inequality, one has that

$$2\mathbf{TV}(f, \phi_1) \leqslant \sqrt{2D(F\|Z)}. \tag{1.5}$$

Our aim is to deduce a bound on $D(F\|Z)$ that is expressed in terms of the so-called *Stein factor* associated with F. Whenever it exists, such a factor is a mapping $\tau_F : \mathbb{R} \to \mathbb{R}$ that is uniquely determined (up to negligible sets) by requiring that $\tau_F(F) \in L^1$ and

$$E[Fg(F)] = E[\tau_F(F) g'(F)]$$

for every smooth test function g. Specifying $g(x) = x$ implies, in particular, that $E[\tau_F(F)] = E[F^2] = 1$. It is easily seen that, under standard regularity assumptions, a version of τ_F is given by $\tau_F(x) = (f(x))^{-1} \int_x^\infty z f(z) dz$, for x in the support of f (in particular, the Stein factor of Z is 1). The relevance of the factor τ_F in comparing F with Z is actually revealed by the following *Stein's bound* [15, 33], which is one of the staples of Stein's method:

$$\mathbf{TV}(f, \phi_1) = \sup \left| E[g'(F)] - E[Fg(F)] \right|, \tag{1.6}$$

where the supremum runs over all continuously differentiable functions $g : \mathbb{R} \to \mathbb{R}$ satisfying $\|g\|_\infty \leqslant \sqrt{2/\pi}$ and $\|g'\|_\infty \leqslant 2$. In particular, from (1.6) one recovers the bound in total variation

$$\mathbf{TV}(f, \phi_1) \leqslant 2\, E[|1 - \tau_F(F)|], \tag{1.7}$$

providing a formal meaning to the intuitive fact that the distributions of F and Z are close whenever τ_F is close to τ_Z, that is, whenever τ_F is close to 1. To motivate the reader, we shall now present a simple illustration of how the estimate (1.7) applies to the usual CLT.

Example 1.2 *Let $\{F_i : i \geqslant 1\}$ be a sequence of i.i.d. copies of F, set $S_n = n^{-1/2} \sum_{i=1}^n F_i$ and assume that $E[\tau_F(F)^2] < +\infty$ (a simple sufficient condition for this to hold is e.g. that f has compact support, and f is bounded from below inside its support). Then, using e.g. [47, Lemma 2],*

$$\tau_{S_n}(S_n) = \frac{1}{n} E\left[\sum_{i=1}^n \tau_F(F_i) \,\Big|\, S_n \right]. \tag{1.8}$$

Since (by definition) $E\left[\tau_{S_n}(S_n)\right] = E\left[\tau_F(F_i)\right] = 1$ for all $i = 1, \ldots, n$ we get

$$\begin{aligned} E\left[(1 - \tau_{S_n}(S_n))^2\right] &= E\left[\left(\frac{1}{n} E\left[\sum_{i=1}^n (1 - \tau_F(F_i)) \,\Big|\, F \right] \right)^2 \right] \\ &\leq \frac{1}{n^2} E\left[\left(\sum_{i=1}^n (1 - \tau_F(F_i)) \right)^2 \right] \\ &= \frac{1}{n^2} \mathrm{Var}\left(\sum_{i=1}^n (1 - \tau_F(F_i)) \right) = \frac{E\left[(1 - \tau_F(F))^2\right]}{n}. \end{aligned} \tag{1.9}$$

In particular, writing f_n for the density of S_n, we deduce from (1.7) *that*

$$\mathbf{TV}(f_n, \phi_1) \leq 2 \frac{\mathrm{Var}(\tau_F(F))^{1/2}}{\sqrt{n}}. \tag{1.10}$$

We shall demonstrate in Section 3 and Section 4 that the quantity $E[|1 - \tau_F(F)|]$ (as well as its multidimensional generalisations) can be explicitly controlled whenever F is a smooth functional of a Gaussian field. In view of these observations, the following question is therefore natural: can one bound $D(F\|Z)$ by an expression analogous to the right-hand-side of (1.7)?

Our strategy for connecting $\tau_F(F)$ and $D(F\|Z)$ is based on an integral version of the classical *de Bruijn's formula* of information theory. To introduce this result, for $t \in [0, 1]$ denote by f_t the density of $F_t = \sqrt{t}F + \sqrt{1-t}Z$, in such a way that $f_1 = f$ and $f_0 = \phi_1$.

Of course $f_t(x) = E\left[\phi_1\left((x - \sqrt{t}F)/\sqrt{1-t}\right)\right]/\sqrt{1-t}$ has support $\mathbb{R}$ and is C^∞ for all $t < 1$. We shall denote by $\rho_t = (\log f_t)'$ the *score function* of F_t (which is, by virtue of the preceding remark, well defined at all $t < 1$ irrespective of the properties of F). For every $t < 1$, the mapping ρ_t is completely characterised by the fact that

$$E[g'(F_t)] = -E[g(F_t)\rho_t(F_t)] \tag{1.11}$$

for every smooth test function g. We also write, for $t \in [0, 1)$,

$$J(F_t) = E[\rho_t(F_t)^2] = \int_{\mathbb{R}} \frac{f_t'(x)^2}{f_t(x)} dx$$

for the *Fisher information* of F_t, and we observe that

$$0 \leqslant E[(F_t + \rho_t(F_t))^2] = J(F_t) - 1 =: J_{st}(F_t),$$

where $J_{st}(F_t)$ is the so-called *standardised Fisher information* of F_t (note that $J_{st}(F_0) = J_{st}(Z) = 0$). With this notation in mind, de Bruijn's formula (in an integral and rescaled version due to Barron [8]) reads

$$D(F\|Z) = \int_0^1 \frac{J(F_t) - 1}{2t} dt = \int_0^1 \frac{J_{st}(F_t)}{2t} dt \tag{1.12}$$

(see Lemma 2.3 below for a multidimensional statement).

Remark 1.3 *Using the standard relation* $J_{st}(F_t) \leqslant tJ_{st}(F) + (1-t)J_{st}(Z) = tJ_{st}(F)$ *(see e.g. [19, Lemma 1.21]), we deduce the upper bound*

$$D(F\|Z) \leqslant \frac{1}{2} J_{st}(F), \tag{1.13}$$

a result which is often proved by using entropy power inequalities (see also Shimizu [46]). Formula (1.13) *is a quantitative counterpart to the intuitive fact that the distributions of F and Z are close, whenever $J_{st}(F)$ is close to zero. Using* (1.5) *we further deduce that closeness between the Fisher informations of F and Z (i.e. $J_{st}(F) \approx 0$) or between the entropies of F and Z (i.e. $D(F\|Z) \approx 0$) both imply closeness in terms of the total variation distance, and hence in terms of many more probability metrics. This observation lies at the heart of the approach from [8, 10, 20] where a fine analysis of the behavior of $\rho_F(F)$ over convolutions (through projection inequalities in the spirit of* (1.8)*) is used to provide explicit bounds on the Fisher information distance which in turn are transformed, by means of de Bruijn's identity* (1.12)*, into bounds on the relative entropy. We will see in Section 4 that the bound* (1.13) *is too crude to be of use in the applications we are interested in.*

Our key result in dimension 1 is the following statement (see Theorem 2.10 for a general multidimensional version), providing a new representation of relative entropy in terms of Stein factors. From now on, we denote by C_c^1 the class of all functions $g : \mathbb{R} \to \mathbb{R}$ that are continuously differentiable and with compact support.

Proposition 1.4 *Let the previous notation prevail. We have*

$$D(F\|Z) = \frac{1}{2} \int_0^1 \frac{t}{1-t} E\big[E[Z(1 - \tau_F(F))|F_t]^2\big] dt. \tag{1.14}$$

Proof. Using $\rho_Z(Z) = -Z$ we see that, for any function $g \in C_c^1$, one has

$$
\begin{aligned}
&E[Z(1-\tau_F(F))g(\sqrt{t}F + \sqrt{1-t}Z)] \\
&= \sqrt{1-t}\, E[(1-\tau_F(F))g'(\sqrt{t}F + \sqrt{1-t}Z)] \\
&= \sqrt{1-t}\big\{E[g'(F_t)] - \frac{1}{\sqrt{t}}E[Fg(F_t)]\big\} \\
&= \frac{\sqrt{1-t}}{t}\big\{E[g'(F_t)] - E[F_t g(F_t)]\big\} = -\frac{\sqrt{1-t}}{t}\, E[(\rho_t(F_t) + F_t)g(F_t)],
\end{aligned}
$$

yielding the representation

$$
\rho_t(F_t) + F_t = -\frac{t}{\sqrt{1-t}} E[Z(1-\tau_F(F))|F_t]. \tag{1.15}
$$

This implies

$$
J(F_t) - 1 = E[(\rho_t(F_t) + F_t)^2] = \frac{t^2}{1-t} E\big[E[Z(1-\tau_F(F))|F_t]^2\big],
$$

and the desired conclusion follows from de Bruijn's identity (1.12). ■

To properly control the integral on the right-hand-side of (1.14), we need to deal with the fact that the mapping $t \mapsto \frac{t}{1-t}$ is not integrable in $t = 1$, so that we cannot directly apply the estimate $E\big[E[Z(1-\tau_F(F))|F_t]^2\big] \leqslant \mathrm{Var}(\tau_F(F))$ to deduce the desired bound. Intuitively, one has to exploit the fact that the mapping $t \mapsto E[Z(1-\tau_F(F))|F_t]$ satisfies $E[Z(1-\tau_F(F))|F_1] = 0$, thus in principle compensating for the singularity at $t \approx 1$.

As we will see below, one can make this heuristic precise provided there exist three constants $c, \delta, \eta > 0$ such that

$$
E[|\tau_F(F)|^{2+\eta}] < \infty \text{ and } E\big[|E[Z(1-\tau_F(F))|F_t]|\big] \leqslant c\, t^{-1}(1-t)^{\delta}, \quad 0 < t \leqslant 1. \tag{1.16}
$$

Under the assumptions appearing in condition (1.16), the following strategy can indeed be implemented in order to deduce a satisfactory bound. First split the integral in two parts: for every $0 < \varepsilon \leqslant 1$,

$$
\begin{aligned}
&2\, D(F\|Z) \\
&\leqslant E[(1-\tau_F(F))^2] \int_0^{1-\varepsilon} \frac{t\,dt}{1-t} + \int_{1-\varepsilon}^1 \frac{t}{1-t}\, E\big[E[Z(1-\tau_F(F))|F_t]^2\big]dt \\
&\leqslant E[(1-\tau_F(F))^2]\, |\log\varepsilon| + \int_{1-\varepsilon}^1 \frac{t}{1-t}\, E\big[E[Z(1-\tau_F(F))|F_t]^2\big]dt,
\end{aligned} \tag{1.17}
$$

the last inequality being a consequence of $\int_0^{1-\varepsilon} \frac{t\,dt}{1-t} = \int_\varepsilon^1 \frac{(1-u)du}{u} \leqslant \int_\varepsilon^1 \frac{du}{u} = -\log\varepsilon$. To deal with the second term in (1.17), let us observe that, by using in particular the Hölder inequality and the convexity of the function $x \mapsto |x|^{\eta+2}$, one deduces from (1.16) that

$$
\begin{aligned}
&E\big[E[Z(1-\tau_F(F))|F_t]^2\big] \\
&= E\left[|E[Z(1-\tau_F(F))|F_t]|^{\frac{\eta}{\eta+1}} |E[Z(1-\tau_F(F))|F_t]|^{\frac{\eta+2}{\eta+1}}\right] \\
&\leqslant E\big[|E[Z(1-\tau_F(F))|F_t]|\big]^{\frac{\eta}{\eta+1}} \times E\big[|E[Z(1-\tau_F(F))|F_t]|^{\eta+2}\big]^{\frac{1}{\eta+1}} \\
&\leqslant c^{\frac{\eta}{\eta+1}}\, t^{-\frac{\eta}{\eta+1}} (1-t)^{\frac{\delta\eta}{\eta+1}} \times E\big[|Z|^{\eta+2}\big]^{\frac{1}{\eta+1}} \times E\big[|1-\tau_F(F)|^{\eta+2}\big]^{\frac{1}{\eta+1}} \\
&\leqslant c^{\frac{\eta}{\eta+1}}\, t^{-1}(1-t)^{\frac{\delta\eta}{\eta+1}} \times 2\, E\big[|Z|^{\eta+2}\big]^{\frac{1}{\eta+1}} \big(1 + E\big[|\tau_F(F)|^{\eta+2}\big]\big)^{\frac{1}{\eta+1}} \\
&= C_\eta\, t^{-1}(1-t)^{\frac{\delta\eta}{\eta+1}},
\end{aligned} \tag{1.18}
$$

with

$$C_\eta := 2c^{\frac{\eta}{\eta+1}} E\big[|Z|^{\eta+2}\big]^{\frac{1}{\eta+1}} \left(1 + E\big[|\tau_F(F)|^{\eta+2}\big]\right)^{\frac{1}{\eta+1}}.$$

By virtue of (1.17) and (1.18), the term $D(F\|Z)$ is eventually amenable to analysis, and one obtains:

$$\begin{aligned} 2\,D(F\|Z) &\leqslant E[(1-\tau_F(F))^2]\,|\log\varepsilon| + C_\eta \int_{1-\varepsilon}^{1} (1-t)^{\frac{\delta\eta}{\eta+1}-1} dt \\ &= E[(1-\tau_F(F))^2]\,|\log\varepsilon| + \frac{C_\eta(\eta+1)}{\delta\eta}\,\varepsilon^{\frac{\delta\eta}{\eta+1}}. \end{aligned}$$

Assuming finally that $E[(1-\tau_F(F))^2] \leqslant 1$ (recall that, in the applications we are interested in, such a quantity is meant to be close to 0) we can optimize over ε and choose $\varepsilon = E[(1-\tau_F(F))^2]^{\frac{\eta+1}{\delta\eta}}$, which leads to

$$\begin{aligned} D(F\|Z) \leqslant& \frac{\eta+1}{2\delta\eta}\, E[(1-\tau_F(F))^2]\,|\log E[(1-\tau_F(F))^2]| \\ &+ \frac{C_\eta(\eta+1)}{2\delta\eta}\, E[(1-\tau_F(F))^2]. \end{aligned} \tag{1.19}$$

Clearly, combining (1.19) with (1.5), one also obtains an estimate in total variation which agrees with (1.7) up to the square root of a logarithmic factor.

The problem is now how to identify sufficient conditions on the law of F for (1.16) to hold; we shall address this issue by means of two auxiliary results. We start with a useful technical lemma, that has been suggested to us by Guillaume Poly.

Lemma 1.5 *Let X be an integrable random variable and let Y be a $\mathbb{R}^d$-valued random vector having an absolutely continuous distribution. Then*

$$E\,|E\,[X\,|\,Y]| = \sup E\,[Xg(Y)]\,, \tag{1.20}$$

where the supremum is taken over all $g \in C_c^1$ such that $\|g\|_\infty \leq 1$.

Proof. Since $|\mathrm{sign}(E[X|Y])| = 1$ we have, by using e.g. Lusin's Theorem,

$$E\,|E\,[X\,|\,Y]| = E[X\mathrm{sign}(E[X|Y])] \leqslant \sup E(Xg(Y)).$$

To see the reversed inequality, observe that, for any g bounded by 1,

$$|E(Xg(Y))| = |E(E(X|Y)g(Y))| \leqslant E\,|E\,[X\,|\,Y]|\,.$$

The lemma is proved. ■

Our next statement relates (1.16) to the problem of estimating the total variation distance between F and $\sqrt{t}F + \sqrt{1-t}x$ for any $x \in \mathbb{R}$ and $0 < t \leqslant 1$.

Lemma 1.6 *Assume that, for some $\kappa, \alpha > 0$,*

$$\mathbf{TV}(\sqrt{t}F + \sqrt{1-t}\,x, F) \leqslant \kappa(1+|x|)t^{-1}(1-t)^\alpha, \quad x \in \mathbb{R},\ t \in (0,1]. \tag{1.21}$$

Then (1.16) holds, with $\delta = \frac{1}{2} \wedge \alpha$ and $c = 4(\kappa+1)$.

Proof. Take $g \in C_c^1$ such that $\|g\|_\infty \leqslant 1$. Then, by independence of Z and F,

$$\begin{aligned}
E\left[Z(1-\tau_F(F))g(F_t)\right] &= E\left[g(F_t)Z\right] - E\left[Zg(F_t)\tau_F(F)\right] \\
&= E\left[g(F_t)Z\right] - \sqrt{1-t}E\left[\tau_F(F)g'(F_t)\right] \\
&= E\left[Z(g(F_t)-g(F))\right] - \sqrt{\frac{1-t}{t}}E\left[g(F_t)F\right]
\end{aligned}$$

so that, since $\|g\|_\infty \leqslant 1$ and $E|F| \leqslant \sqrt{E[F^2]} = 1$,

$$\begin{aligned}
\left|E\left[Z(1-\tau_F(F))g(F_t)\right]\right| &\leqslant \left|E\left[Z\left(g(F_t)-g(F)\right)\right]\right| + \sqrt{\frac{1-t}{t}} \\
&\leqslant \left|E\left[Z\left(g(F_t)-g(F)\right)\right]\right| + t^{-1}\sqrt{1-t}.
\end{aligned}$$

We have furthermore

$$\begin{aligned}
\left|E\left[Z(g(F_t)-g(F))\right]\right| &= \left|\int_{\mathbb{R}} xE[g(\sqrt{t}F+\sqrt{1-t}\,x)-g(F)]\phi_1(x)dx\right| \\
&\leqslant 2\int_{\mathbb{R}} |x|\ \mathbf{TV}(\sqrt{t}F+\sqrt{1-t}\,x, F)\phi_1(x)dx \\
&\leqslant 2\kappa t^{-1}(1-t)^\alpha \int_{\mathbb{R}} |x|(1+|x|)\phi_1(x)dx \\
&\leqslant 4\kappa\, t^{-1}\,(1-t)^\alpha.
\end{aligned}$$

Inequality (1.16) now follows by applying Lemma 1.5. ■

As anticipated, in Section 4 (see Lemma 4.4 for a precise statement) we will describe a wide class of distributions satisfying (1.21). The previous discussion yields finally the following statement, answering the original question of providing a bound on $D(F\|Z)$ that is comparable with the estimate (1.7).

Theorem 1.7 *Let F be a random variable with density $f:\mathbb{R}\to[0,\infty)$, satisfying $E[F]=0$ and $E[F^2]=1$. Let $Z\sim\mathscr{N}(0,1)$ be a standard Gaussian variable (independent of F). If, for some $\alpha,\kappa,\eta>0$, one has*

$$E[|\tau_F(F)|^{2+\eta}] < \infty \tag{1.22}$$

and

$$\mathbf{TV}(\sqrt{t}F+\sqrt{1-t}\,x, F) \leqslant \kappa(1+|x|)t^{-1}(1-t)^\alpha, \quad x\in\mathbb{R},\ t\in(0,1], \tag{1.23}$$

then, provided $\Delta := E[(1-\tau_F(F))^2] \leqslant 1$,

$$D(F\|Z) \leqslant \frac{\eta+1}{(1\wedge 2\alpha)\eta}\,\Delta\,|\log\Delta| + \frac{C_\eta(\eta+1)}{(1\wedge 2\alpha)\eta}\,\Delta, \tag{1.24}$$

where

$$C_\eta = 2(4\kappa+4)^{\frac{\eta}{\eta+1}} E\left[|Z|^{\eta+2}\right]^{\frac{1}{\eta+1}} \left(1+E\left[|\tau_F(F)|^{\eta+2}\right]\right)^{\frac{1}{\eta+1}}.$$

1.3 Plan

The rest of the paper is organised as follows. In Section 2 we will prove that Theorem 1.7 can be generalised to a fully multidimensional setting. Section 3 contains some general results related to (infinite-dimensional) Gaussian stochastic analysis. Finally, in Section 4 we shall apply our estimates in order to deduce general bounds of the type appearing in Theorem 1.1.

2 Entropy bounds via de Bruijn's identity and Stein matrices

In Section 2.1 and Section 2.2 we discuss some preliminary notions related to the theory of information (definitions, notations and main properties). Section 2.3 contains the proof of a new integral formula, allowing one to represent the relative entropy of a given random vector in terms of a Stein matrix. The reader is referred to the monograph [19], as well as to [1, Chapter 10], for any unexplained definition and result concerning information theory.

2.1 Entropy

Fix an integer $d \geqslant 1$. Throughout this section, we consider a d-dimensional square-integrable and centered random vector $F = (F_1, ..., F_d)$ with covariance matrix $B > 0$. We shall assume that the law of F admits a density $f = f_F$ (with respect to the Lebesgue measure) with support $S \subseteq \mathbb{R}^d$. No other assumptions on the distribution of F will be needed. We recall that the *differential entropy* (or, simply, the *entropy*) of F is given by the quantity $\mathrm{Ent}(F) := -E[\log f(F)] = -\int_{\mathbb{R}^d} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = -\int_S f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}$, where we have adopted (here and for the rest of the paper) the standard convention $0 \log 0 := 0$. Note that $\mathrm{Ent}(F) = \mathrm{Ent}(F + c)$ for all $c \in \mathbb{R}^d$, i.e. entropy is location invariant.

As discussed above, we are interested in estimating the distance between the law of F and the law of a d-dimensional centered Gaussian vector $Z = (Z_1, ..., Z_d) \sim \mathscr{N}_d(0, C)$, where $C > 0$ is the associated covariance matrix. Our measure of the discrepancy between the distributions of F and Z is the *relative entropy* (often called *Kullback-Leibler divergence* or *information entropy*)

$$D(F||Z) := E\left[\log(f(F)/\phi(F))\right] = \int_{\mathbb{R}^d} f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{\phi(\mathbf{x})}\right) d\mathbf{x}, \tag{2.25}$$

where $\phi = \phi_d(\cdot\,; C)$ is the density of Z. It is easy to compute the Gaussian entropy $\mathrm{Ent}(Z) = 1/2 \log\left((2\pi e)^d |C|\right)$ (where $|C|$ is the determinant of C), from which we deduce the following alternative expression for the relative entropy

$$0 \leqslant D(F||Z) = \mathrm{Ent}(Z) - \mathrm{Ent}(F) + \frac{\mathrm{tr}(C^{-1}B) - d}{2}, \tag{2.26}$$

where 'tr' stands for the usual trace operator. If Z and F have the same covariance matrix then the relative entropy is simply the entropy gap between F and Z so that, in particular, one infers from (2.26) that Z has maximal entropy among all absolutely continuous random vectors with covariance matrix C.

We stress that the relative entropy D does not define a *bona fide* probability distance (for absence of a triangle inequality, as well as for lack of symmetry): however, one can easily translate estimates on the relative entropy in terms of the total variation distance, using the already recalled Pinsker-Csiszar-Kullback inequality (1.3). In the next subsection, we show how one can represent the quantity $D(F||Z)$ as the integral of the standardized Fisher information of some adequate interpolation between F and Z.

2.2 Fisher information and de Bruijn's identity

Without loss of generality, we may assume for the rest of the paper that the vectors F and Z (as defined in the previous Section 2.1) are stochastically independent. For every $t \in [0, 1]$, we define the centered random vector $F_t := \sqrt{t}F + \sqrt{1-t}Z$, in such a way that $F_0 = Z$ and $F_1 = F$. It is clear that F_t is centered and has covariance $\Gamma_t = tB + (1-t)C > 0$; moreover, whenever $t \in [0, 1)$, F_t has a strictly positive and

infinitely differentiable density, that we shall denote by f_t (see e.g. [21, Lemma 3.1] for more details). For every $t \in [0, 1)$, we define the *score* of F_t as the $\mathbb{R}^d$-valued function given by

$$\rho_t : \mathbb{R}^d \to \mathbb{R}^d : \mathbf{x} \mapsto \rho_t(\mathbf{x}) = (\rho_{t,1}(\mathbf{x}), ..., \rho_{t,d}(\mathbf{x}))^T := \nabla \log f_t(\mathbf{x}), \tag{2.27}$$

with ∇ the usual gradient in $\mathbb{R}^d$ (note that we will systematically regard the elements of $\mathbb{R}^d$ as column vectors). The quantity $\rho_t(\mathbf{x})$ is of course well-defined for every $\mathbf{x} \in \mathbb{R}^d$ and every $t \in [0, 1)$; moreover, it is easily seen that the random vector $\rho_t(F_t)$ is completely characterized (up to sets of P-measure zero) by the relation

$$E[\rho_t(F_t)g(F_t)] = -E[\nabla g(F_t)], \tag{2.28}$$

holding for every smooth function $g : \mathbb{R}^d \to \mathbb{R}$. Selecting $g = 1$ in (2.28), one sees that $\rho_t(F_t)$ is a centered random vector. The covariance matrix of $\rho_t(F_t)$ is denoted by

$$J(F_t) := E[\rho_t(F_t)\rho_t(F_t)^T] \tag{2.29}$$

(with components $J(F_t)_{ij} = E[\rho_{t,i}(F_t)\rho_{t,j}(F_t)]$ for $1 \le i, j \le d$), and is customarily called the *Fisher information matrix* of F_t. Focussing on the case $t = 0$, one sees immediately that the Gaussian vector $F_0 = Z \sim \mathscr{N}_d(0, C)$ has linear score function $\rho_0(\mathbf{x}) = \rho_Z(\mathbf{x}) = -C^{-1}\mathbf{x}$ and Fisher information $J(F_0) = J(Z) = C^{-1}$.

Remark 2.1 *Fix $t \in [0, 1)$. Using formula (2.28) one deduces that a version of $\rho_t(F_t)$ is given by the conditional expectation $-(1-t)^{-1/2}E[C^{-1}Z|F_t]$, from which we infer that the matrix $J(F_t)$ is well-defined and its entries are all finite.*

For $t \in [0, 1)$, we define the *standardized Fisher information matrix* of F_t as

$$J_{st}(F_t) := \Gamma_t E\left[\left(\rho_t(F_t) + \Gamma_t^{-1}F_t\right)\left(\rho_t(F_t) + \Gamma_t^{-1}F_t\right)^T\right] = \Gamma_t J(F_t) - I_d, \tag{2.30}$$

where I_d is the $d \times d$ identity matrix, and the last equality holds because $E[\rho_t(F_t)F_t] = -I_d$. Note that the positive semidefinite matrix $\Gamma_t^{-1}J_{st}(F_t) = J(F_t) - \Gamma_t^{-1}$ is the difference between the Fisher information matrix of F_t and that of a Gaussian vector having distribution $\mathscr{N}_d(0, \Gamma_t)$. Observe that

$$J_{st}(F_t) := E\left[\left(\rho_t^\star(F_t) + F_t\right)\left(\rho_t^\star(F_t) + F_t\right)^T\right]\Gamma_t^{-1}, \tag{2.31}$$

where the vector

$$\rho_t^\star(F_t) = (\rho_{t,1}^\star(F_t), ..., \rho_{t,d}^\star(F_t))^T := \Gamma_t \rho_t(F_t) \tag{2.32}$$

is completely characterized (up to sets of P-measure 0) by the equation

$$E[\rho_t^\star(F_t)g(F_t)] = -\Gamma_t E[\nabla g(F_t)], \tag{2.33}$$

holding for every smooth test function g.

Remark 2.2 *Of course the above information theoretic quantities are not defined only for Gaussian mixtures of the form F_t but more generally for any random vector satisfying the relevant assumptions (which are necessarily verified by the F_t). In particular, if F has covariance matrix B and differentiable density f then, letting $\rho_F(\mathbf{x}) := \nabla \log f(\mathbf{x})$ be the score function for F, the standardized Fisher information of F is*

$$J_{st}(F) = BE\left[\rho_F(F)\rho_F(F)^T\right]. \tag{2.34}$$

In the event that the above be well-defined then it is also scale invariant in the sense that $J_{st}(\alpha F) = J_{st}(F)$ for all $\alpha \in \mathbb{R}$.

The following fundamental result is known as the (multidimensional) *de Bruijn's identity*: it shows that the relative entropy $D(F||Z)$ can be represented in terms of the integral of the mapping $t \mapsto \mathrm{tr}(C\Gamma_t^{-1}J_{st}(F_t))$ with respect to the measure $dt/2t$ on $(0,1]$. It is one of the staples of the entire paper. We refer the reader e.g. to [1, 8] for proofs in the case $d = 1$. Our multidimensional statement is a rescaling of [21, Theorem 2.3] (some more details are given in the proof).

Lemma 2.3 (Multivariate de Bruijn's identity) *Let the above notation and assumptions prevail. Then,*

$$D(F||Z) = \int_0^1 \frac{1}{2t}\mathrm{tr}\left(C\Gamma_t^{-1}J_{st}(F_t)\right)dt \qquad (2.35)$$
$$+\frac{1}{2}\left(\mathrm{tr}\left(C^{-1}B\right)-d\right)+\int_0^1 \frac{1}{2t}\mathrm{tr}\left(C\Gamma_t^{-1}-I_d\right)dt.$$

Proof. In [21, Theorem 2.3] it is proved that

$$D(F||Z) = \int_0^\infty \frac{1}{2}\mathrm{tr}\left(C(B+\tau C)^{-1}J_{st}(F+\sqrt{\tau}Z)\right)d\tau$$
$$+\frac{1}{2}\left(\mathrm{tr}\left(C^{-1}B\right)-d\right)+\frac{1}{2}\int_0^\infty \mathrm{tr}\left(C\left((B+\tau C)^{-1}-\frac{C^{-1}}{1+\tau}\right)\right)d\tau$$

(note that the definition of standardized Fisher information used in [21] is different from ours). The conclusion is obtained by using the change of variables $t = (1+\tau)^{-1}$, as well as the fact that

$$J_{st}\left(F+\sqrt{\frac{1-t}{t}}Z\right) = J_{st}\left(\sqrt{t}F+\sqrt{1-t}Z\right),$$

which follows from the scale-invariance of standardized Fisher information mentionned in Remark 2.2. ■

Remark 2.4 *Assume that C_n, $n \geqslant 1$, is a sequence of $d \times d$ nonsingular covariance matrices such that $C_{n;i,j} \to B_{i,j}$ for every $i,j = 1,...,d$, as $n \to \infty$. Then, the second and third summands of* (2.35) *(with C_n replacing C) converge to 0 as $n \to \infty$.*

For future reference, we will now rewrite formula (2.35) for some specific choices of d, F, B and C.

Example 2.5 *(i) Assume $F \sim \mathscr{N}_d(0,B)$. Then, $J_{st}(F_t) = 0$ (null matrix) for every $t \in [0,1)$, and formula (2.35) becomes*

$$D(F||Z) = \frac{1}{2}\left(\mathrm{tr}\left(C^{-1}B\right)-d\right)+\int_0^1 \frac{1}{2t}\mathrm{tr}\left(C\Gamma_t^{-1}-I_d\right)dt. \qquad (2.36)$$

(ii) Assume that $d = 1$ and that F and Z have variances $b, c > 0$, respectively. Defining $\gamma_t = tb+(1-t)c$, relation (2.35) becomes

$$D(F||Z) = \int_0^1 \frac{c}{2t\gamma_t}J_{st}(F_t)dt + \frac{1}{2}\left(\frac{b}{c}-1\right)+\int_0^1 \frac{1}{2t}\left(\frac{c}{\gamma_t}-1\right)dt \qquad (2.37)$$
$$= \int_0^1 \frac{c}{2t}E[(\rho_t(F_t)+\gamma_t^{-1}F_t)^2]dt + \frac{1}{2}\left(\frac{b}{c}-1\right)+\frac{\log c - \log b}{2}.$$

Relation (2.37) in the case $b = c$ $(= \gamma_t)$ corresponds to the integral formula (1.12) *proved by Barron in [8, Lemma 1].*

(iii) If $B = C$, then (2.35) takes the form

$$D(F||Z) = \int_0^1 \frac{1}{2t} \mathrm{tr}\,(J_{st}(F_t))\, dt. \tag{2.38}$$

In the special case where $B = C = I_d$, one has that

$$D(F||Z) = \frac{1}{2} \sum_{j=1}^{d} \int_0^1 \frac{1}{t} E\left[(\rho_{t,j}(F_t) + F_{t,j})^2\right] dt, \tag{2.39}$$

of which (1.12) *is a particular case ($d = 1$).*

In the univariate setting, the general variance case ($E\left[F^2\right] = \sigma^2$) follows trivially from the standardized one ($E\left[F^2\right] = 1$) through scaling; the same cannot be said in the multivariate setting since the appealing form (2.39) cannot be directly achieved for $d \geq 2$ when the covariance matrices are not the identity because here the dependence structure of F needs to be taken into account. In Lemma 2.6 we provide an estimate allowing one to deal with this difficulty in the case $B = C$, for every d. The proof is based on the following elementary fact: if A, B are two $d \times d$ symmetric matrices, and if A is semi-positive definite, then

$$\lambda_{\min}(B) \times \mathrm{tr}(A) \leqslant \mathrm{tr}(AB) \leqslant \lambda_{\max}(B) \times \mathrm{tr}(A), \tag{2.40}$$

where $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ stand, respectively, for the maximum and minimum eigenvalue of B. Observe that $\lambda_{\max}(B) = \|B\|_{op}$, the operator norm of B.

Lemma 2.6 *Fix $d \geqslant 1$, and assume that $B = C$. Then, $C\Gamma_t^{-1} = I_d$, and one has the following estimates*

$$\lambda_{\min}(C) \times \sum_{j=1}^{d} E[(\rho_{t,j}(F_t) + (C^{-1}F_t)_j)^2] \leqslant \mathrm{tr}\,(J_{st}(F_t)) \tag{2.41}$$

$$\leqslant \lambda_{\max}(C) \times \sum_{j=1}^{d} E[(\rho_{t,j}(F_t) + (C^{-1}F_t)_j)^2],$$

$$\lambda_{\min}(C^{-1}) \times \sum_{j=1}^{d} E[(\rho^{\star}_{t,j}(F_t) + F_{t,j})^2] \leqslant \mathrm{tr}\,(J_{st}(F_t)) \tag{2.42}$$

$$\leqslant \lambda_{\max}(C^{-1}) \times \sum_{j=1}^{d} E[(\rho^{\star}_{t,j}(F_t) + F_{t,j})^2].$$

Proof. Write $\mathrm{tr}\,(J_{st}(F_t)) = \mathrm{tr}\left(C^{-1}J_{st}(F_t)C\right)$ and apply (2.40) first to $A = C^{-1}J_{st}(F_t)$ and $B = C$, and then to $A = J_{st}(F_t)C$ and $B = C^{-1}$. ■

In the next section, we prove a new representation of the quantity $\rho_t(F_t) + C^{-1}F_t$ in terms of Stein matrices: this connection will provide the ideal framework in order to deal with the normal approximation of general random vectors.

2.3 Stein matrices and a key lemma

The centered d-dimensional vectors F, Z are defined as in the previous section (in particular, they are stochastically independent).

Definition 2.7 (Stein matrices) Let $M(d, \mathbb{R})$ denote the space of $d \times d$ real matrices. We say that the matrix-valued mapping

$$\tau_F : \mathbb{R}^d \to M(d, \mathbb{R}) : \mathbf{x} \mapsto \tau_F(\mathbf{x}) = \{\tau_F^{i,j}(\mathbf{x}) : i, j = 1, ..., d\}$$

is a *Stein matrix* for F if $\tau_F^{i,j}(F) \in L^1$ for every i, j and the following equality is verified for every differentiable function $g : \mathbb{R}^d \to \mathbb{R}$ such that both sides are well-defined:

$$E\left[Fg(F)\right] = E\left[\tau_F(F)\nabla g(F)\right], \tag{2.43}$$

or, equivalently,

$$E\left[F_i g(F)\right] = \sum_{j=1}^{d} E\left[\tau_F^{i,j}(F)\partial_j g(F)\right], \quad i = 1, ..., d. \tag{2.44}$$

The entries of the random matrix $\tau_F(F)$ are called the *Stein factors* of F.

Remark 2.8 *(i) Selecting $g(F) = F_j$ in (2.43), one deduces that, if τ_F is a Stein matrix for F, then $E[\tau_F(F)] = C$. More to this point, if $F \sim \mathscr{N}_d(0, C)$, then the covariance matrix C is itself a Stein matrix for F. This last relation is known as the Stein's identity for the multivariate Gaussian distribution.*

(ii) Assume that $d = 1$ and that F has density f and variance $b > 0$. Then, under some standard regularity assumptions, it is easy to see that $\tau_F(x) = b \int_x^\infty y f(y)dy/f(x)$ is a Stein factor for F.

Lemma 2.9 (Key Lemma) *Let the above notation and framework prevail, and assume that τ_F is a Stein matrix for F such that $\tau_F^{i,j}(F) \in L^1(\Omega)$ for every $i, j = 1, ..., d$. Then, for every $t \in [0, 1)$, the mapping*

$$\mathbf{x} \mapsto -\frac{t}{\sqrt{1-t}} E\Big[\big(I_d - C^{-1}\tau_F(F)\big)C^{-1}Z \;\Big|\; F_t = \mathbf{x}\Big] - C^{-1}\mathbf{x} \tag{2.45}$$

is a version of the score ρ_t of F_t. Also, the mapping

$$\mathbf{x} \mapsto -\frac{t}{\sqrt{1-t}} E\Big[\big(\Gamma_t - \Gamma_t C^{-1}\tau_F(F)\big)C^{-1}Z \;\Big|\; F_t = \mathbf{x}\Big] - \Gamma_t C^{-1}\mathbf{x} \tag{2.46}$$

is a version of the function $\rho_t^\star$ defined in formula (2.32).

Proof. Remember that $-C^{-1}Z$ is the score of Z, and denote by $\mathbf{x} \mapsto A_t(\mathbf{x})$ the mapping defined in (2.45). Removing the conditional expectation and exploiting the independence of F and Z, we infer that, for every smooth test function g,

$$\begin{aligned}
E\left[A_t(F_t)g(F_t)\right] &= \frac{t}{\sqrt{1-t}} E\Big[\big(I_d - C^{-1}\tau_F(F)\big)\left(-C^{-1}Z\right) g(F_t)\Big] \\
&\quad - \sqrt{t} C^{-1} E\left[Fg(F_t)\right] - \sqrt{1-t} E\left[C^{-1}Zg(F_t)\right] \\
&= -tE\left[\big(I_d - C^{-1}\tau_F(F)\big)\nabla g(F_t)\right] \\
&\quad - tC^{-1}E\left[\tau_F(F)\nabla g(F_t)\right] - (1-t)E\left[\nabla g(F_t)\right] \\
&= -E\left[\nabla g(F_t)\right],
\end{aligned}$$

thus yielding the desired conclusion. ∎

To simplify the forthcoming discussion, we shall use the shorthand notation: $\widetilde{Z} = (\widetilde{Z}_d, ..., \widetilde{Z}_d) := C^{-1}Z \sim \mathscr{N}_d(0, C^{-1})$, $\widetilde{F} = (\widetilde{F}_d, ..., \widetilde{F}_d) := C^{-1}F$, and $\widetilde{\tau}_F = \{\widetilde{\tau}_F^{i,j} : i, j = 1, ..., d\} := C^{-1}\tau_F$. The following statement is the main achievement of the section, and is obtained by combining Lemma 2.9 with formulae (2.38) and (2.41)–(2.42), in the case where $C = B$.

Theorem 2.10 *Let the above notation and assumptions prevail, assume that $B = C$, and introduce the notation*

$$A_1(F;Z) := \frac{1}{2}\int_0^1 \frac{t}{1-t} E\left[\sum_{j=1}^d E\left[\sum_{k=1}^d (\mathbf{1}_{j=k} - \widetilde{\tau}_F^{j,k}(F))\widetilde{Z}_k \;\Big|\; F_t\right]^2\right] dt, \tag{2.47}$$

$$A_2(F;Z) := \frac{1}{2}\int_0^1 \frac{t}{1-t} E\left[\sum_{j=1}^d E\left[\sum_{k=1}^d (C(j,k) - \tau_F^{j,k}(F))\widetilde{Z}_k \;\Big|\; F_t\right]^2\right] dt. \tag{2.48}$$

Then one has the inequalities

$$\lambda_{\min}(C) \times A_1(F;Z) \leqslant D(F\|Z) \leqslant \lambda_{\max}(C) \times A_1(F;Z), \tag{2.49}$$

$$\lambda_{\min}(C^{-1}) \times A_2(F;Z) \leqslant D(F\|Z) \leqslant \lambda_{\max}(C^{-1}) \times A_2(F;Z). \tag{2.50}$$

In particular, when $C = B = I_d$,

$$D(F\|Z) = \frac{1}{2}\int_0^1 \frac{t}{1-t} E\left[\sum_{j=1}^d E\left[\sum_{k=1}^d (\mathbf{1}_{j=k} - \tau_F^{j,k}(F))Z_k \;\Big|\; F_t\right]^2\right] dt. \tag{2.51}$$

The next subsection focusses on general bounds based on the estimates (2.47)–(2.51).

2.4 A general bound

The following statement provides the announced multidimensional generalisation of Theorem 1.7. In particular, the main estimate (2.55) provides an explicit quantitative counterpart to the heuristic fact that, if there exists a Stein matrix τ_F such that $\|\tau_F - C\|_{H.S.}$ is small (with $\|\cdot\|_{H.S.}$ denoting the usual Hilbert-Schmidt norm), then the distribution of F and $Z \sim \mathscr{N}_d(0,C)$ must be close. By virtue of Theorem 2.10, the proximity of the two distributions is expressed in terms of the relative entropy $D(F\|Z)$.

Theorem 2.11 *Let F be a centered and square integrable random vector with density $f : \mathbb{R}^d \to [0,\infty)$, let $C > 0$ be its covariance matrix, and assume that τ_F is a Stein matrix for F. Let $Z \sim \mathscr{N}_d(0,C)$ be a Gaussian random vector independent of F. If, for some $\kappa, \eta > 0$ and $\alpha \in (0, \frac{1}{2}]$, one has*

$$E\big[|\tau_F^{j,k}(F)|^{\eta+2}\big] < \infty, \quad j,k = 1,\dots,d, \tag{2.52}$$

as well as

$$\mathbf{TV}\Big(\sqrt{t}F + \sqrt{1-t}\,\mathbf{x}, F\Big) \leqslant \kappa(1+\|\mathbf{x}\|_1)(1-t)^\alpha, \quad \mathbf{x} \in \mathbb{R}^d,\, t \in [1/2, 1], \tag{2.53}$$

then, provided

$$\Delta := E[\|C - \tau_F\|_{H.S.}^2] = \sum_{j,k=1}^d E\left[\Big(C(j,k) - \tau_F^{j,k}(F)\Big)^2\right] \leqslant 2^{-\frac{\eta+1}{\alpha\eta}}, \tag{2.54}$$

one has

$$D(F\|Z) \leqslant \frac{d(\eta+1)\lambda_{\max}(C^{-1})}{2\alpha\eta} \max_{1\leqslant l\leqslant d} E[(\widetilde{Z}_l)^2] \times \Delta\,|\log\Delta| + \frac{C_{d,\eta,\tau}(\eta+1)\lambda_{\max}(C^{-1})}{2\alpha\eta}\Delta, \tag{2.55}$$

where

$$C_{d,\eta,\tau} := 2d^2 \left(2\kappa\, E[\|Z\|_1(1+\|Z\|_1)] + \sqrt{\max_j C(j,j)} \right) \max_{1\leqslant l\leqslant d} E[|\widetilde{Z}_l]^{\eta+2}]^{\frac{1}{\eta+1}}$$
$$\times \sum_{j,k=1}^{d} \left(|C(j,k)|^{\eta+2} + E\left[|\tau_F^{j,k}(F)|^{\eta+2}\right] \right)^{\frac{1}{\eta+1}}, \tag{2.56}$$

and (as above) $\widetilde{Z} = C^{-1}Z$.

Proof. Take $g \in C_c^1$ such that $\|g\|_\infty \leqslant 1$. Then, by independence of $\widetilde{Z}$ and F and using (2.44), one has, for any $j = 1, \dots, d$,

$$\begin{aligned}
&E\left[\sum_{k=1}^{d}(C(j,k) - \tau_F^{j,k}(F))\widetilde{Z}_k\, g(F_t)\right] \\
&= \sum_{k,l=1}^{d} C(j,k)C^{-1}(k,l)\, E\left[Z_l\, g(F_t)\right] - \sum_{k,l=1}^{d} C^{-1}(k,l) E\left[\tau_F^{j,k}(F) Z_l\, g(F_t)\right] \\
&= E\left[Z_j\, g(F_t)\right] - \sqrt{1-t} \sum_{k,l,m=1}^{d} C^{-1}(k,l)C(l,m) E\left[\tau_F^{j,k}(F)\partial_m g(F_t)\right] \\
&= E\left[Z_j\, (g(F_t) - g(F))\right] - \sqrt{1-t}\sum_{k=1}^{d} E\left[\tau_F^{j,k}(F)\partial_k g(F_t)\right] \\
&= E\left[Z_j\, (g(F_t) - g(F))\right] - \sqrt{\frac{1-t}{t}} E\left[F_j g(F_t)\right].
\end{aligned}$$

Using (2.53), we have

$$\begin{aligned}
&|E\left[Z_j(g(F_t) - g(F))\right]| \\
&= \left| \int_{\mathbb{R}^d} x_j E[g(\sqrt{t}F + \sqrt{1-t}\,\mathbf{x}) - g(F)]\, \frac{e^{-\frac{1}{2}\langle C^{-1}\mathbf{x},\mathbf{x}\rangle}}{(2\pi)^{\frac{d}{2}}\sqrt{\det C}}\, d\mathbf{x} \right| \\
&\leqslant 2\int_{\mathbb{R}^d} |x_j| \mathbf{TV}(\sqrt{t}F + \sqrt{1-t}\,\mathbf{x}, F)\, \frac{e^{-\frac{1}{2}\langle C^{-1}\mathbf{x},\mathbf{x}\rangle}}{(2\pi)^{\frac{d}{2}}\sqrt{\det C}}\, d\mathbf{x} \\
&\leqslant 2\kappa(1-t)^\alpha \int_{\mathbb{R}^d} \|\mathbf{x}\|_1(1+\|\mathbf{x}\|_1)\, \frac{e^{-\frac{1}{2}\langle C^{-1}\mathbf{x},\mathbf{x}\rangle}}{(2\pi)^{\frac{d}{2}}\sqrt{\det C}}\, d\mathbf{x} \\
&= 2\kappa\, E[\|Z\|_1(1+\|Z\|_1)]\, (1-t)^\alpha.
\end{aligned}$$

As a result, due to Lemma 1.5 and since $E|F_j| \leqslant \sqrt{E[F_j^2]} \leqslant \sqrt{\max_j C(j,j)}$, one obtains

$$\begin{aligned}
&\max_{1\leqslant j\leqslant d} E\left[\left| E\left[\sum_{k=1}^{d}(C(j,k) - \tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]\right|\right] \\
&\leqslant \left(2\kappa\, E[\|Z\|_1(1+\|Z\|_1)] + \sqrt{\max_j C(j,j)} \right)(1-t)^\alpha.
\end{aligned}$$

Now, using among others the Hölder inequality and the convexity of the function $x \mapsto |x|^{\eta+2}$, we have that

$$
\begin{aligned}
&\sum_{j=1}^{d} E\left[E\left[\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]^2\right] \\
&= \sum_{j=1}^{d} E\left[\left|E\left[\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]\right|^{\frac{\eta}{\eta+1}} \times\right. \\
&\qquad\qquad \left.\times \left|E\left[\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]\right|^{\frac{\eta+2}{\eta+1}}\right] \\
&\leqslant \sum_{j=1}^{d} E\left[\left|E\left[\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]\right|\right]^{\frac{\eta}{\eta+1}} \times \\
&\qquad\qquad \times E\left[\left|E\left[\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]\right|^{\eta+2}\right]^{\frac{1}{\eta+1}} \\
&\leqslant C_{d,\eta,\tau}\,(1-t)^{\frac{\alpha\eta}{\eta+1}},
\end{aligned} \tag{2.57}
$$

with $C_{d,\eta,\tau}$ given by (2.56). At this stage, we shall use the key identity (2.50). To properly control the right-hand-side of (2.48), we split the integral in two parts: for every $0 < \varepsilon \leqslant \frac{1}{2}$,

$$
\begin{aligned}
&\frac{2}{\lambda_{max}(C^{-1})}\, D(F\|Z) \\
&\leqslant \sum_{j=1}^{d} E\left[\left(\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k\right)^2\right] \int_0^{1-\varepsilon} \frac{t\,dt}{1-t} \\
&\quad + \int_{1-\varepsilon}^{1} \frac{t}{1-t} \sum_{j=1}^{d} E\left[E\left[\sum_{k=1}^{d}(C(j,k)-\tau_F^{j,k}(F))\widetilde{Z}_k \middle| F_t\right]^2\right] dt \\
&\leqslant d \max_{1\leqslant l\leqslant d} E[(\widetilde{Z}_l)^2] \sum_{j,k=1}^{d} E\left[\left(C(j,k)-\tau_F^{j,k}(F)\right)^2\right] |\log\varepsilon| \\
&\quad + C_{d,\eta,\tau} \int_{1-\varepsilon}^{1} (1-t)^{\frac{\alpha\eta}{\eta+1}-1} dt \\
&\leqslant d \max_{1\leqslant l\leqslant d} E[(\widetilde{Z}_l)^2] \sum_{j,k=1}^{d} E\left[\left(C(j,k)-\tau_F^{j,k}(F)\right)^2\right] |\log\varepsilon| + C_{d,\eta,\tau}\frac{\eta+1}{\alpha\eta}\,\varepsilon^{\frac{\alpha\eta}{\eta+1}},
\end{aligned}
$$

the second inequality being a consequence of (2.57) as well as $\int_0^{1-\varepsilon}\frac{t\,dt}{1-t} = \int_\varepsilon^1 \frac{(1-u)du}{u} \leqslant \int_\varepsilon^1 \frac{du}{u} = -\log\varepsilon$. Since (2.54) holds true, one can optimize over ε and choose $\varepsilon = \Delta^{\frac{\eta+1}{\alpha\eta}}$, which leads to the desired estimate (2.55). ■

Before applying the content of Theorem 2.11 to entropic CLTs on a Gaussian space (see Section 4), we devote the forthcoming Section 3 to some preliminary results about Gaussian stochastic analysis.

3 Gaussian spaces and variational calculus

As announced, we shall now focus on random variables that can be written as functionals of a countable collection of independent and identically distributed Gaussian $\mathscr{N}(0,1)$ random variables, that we shall denote by

$$\mathbf{G} = \Big\{ G_i : i \geqslant 1 \Big\}. \tag{3.58}$$

Note that our description of $\mathbf{G}$ is equivalent to saying that $\mathbf{G}$ is a Gaussian sequence such that $E[G_i] = 0$ for every i and $E[G_iG_j] = \mathbf{1}_{\{i=j\}}$. We will write $L^2(\sigma(\mathbf{G})) := L^2(P, \sigma(\mathbf{G}))$ to indicate the class of square-integrable (real-valued) random variables that are measurable with respect to the σ-field generated by $\mathbf{G}$.

The reader is referred e.g. to [33, 38] for any unexplained definition or result appearing in the subsequent subsections.

3.1 Wiener chaos

We will now briefly introduce the notion of *Wiener chaos*.

Definition 3.1 (Hermite polynomials and Wiener chaos)

1. *The sequence of Hermite polynomials $\{H_m : m \geqslant 0\}$ is defined as follows: $H_0 = 1$, and, for $m \geqslant 1$,*

$$H_m(x) = (-1)^m e^{\frac{x^2}{2}} \frac{d^m}{dx^m} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

   It is a standard result that the sequence $\{(m!)^{-1/2} H_m : m \geqslant 0\}$ is an orthonormal basis of $L^2(\mathbb{R}, \phi_1(x)dx)$.

2. *A multi-index $\alpha = \{\alpha_i : i \geqslant 1\}$ is a sequence of nonnegative integers such that $\alpha_i \neq 0$ only for a finite number of indices i. We use the symbol Λ in order to indicate the collection of all multi-indices, and use the notation $|\alpha| = \sum_{i\geqslant 1} \alpha_i$, for every $\alpha \in \Lambda$.*

3. *For every integer $q \geqslant 0$, the qth Wiener chaos associated with $\mathbf{G}$ is defined as follows: $C_0 = \mathbb{R}$, and, for $q \geqslant 1$, C_q is the $L^2(P)$-closed vector space generated by random variables of the type*

$$\Phi(\alpha) = \prod_{i=1}^{\infty} H_{\alpha_i}(G_i), \quad \alpha \in \Lambda \text{ and } |\alpha| = q. \tag{3.59}$$

It is easily seen that two random variables belonging to Wiener chaoses of different orders are orthogonal in $L^2(\sigma(\mathbf{G}))$. Moreover, since linear combinations of polynomials are dense in $L^2(\sigma(\mathbf{G}))$, one has that $L^2(\sigma(\mathbf{G})) = \bigoplus_{q\geqslant 0} C_q$, that is, any square-integrable functional of $\mathbf{G}$ can be written as an infinite sum, converging in L^2 and such that the qth summand is an element of C_q. This orthogonal decomposition of $L^2(\sigma(\mathbf{G}))$ is customarily called the *Wiener-Itô chaotic decomposition* of $L^2(\sigma(\mathbf{G}))$.

It is often convenient to encode random variables in the spaces C_q by means of increasing tensor powers of Hilbert spaces. To do this, introduce an (arbitrary) separable real Hilbert space $\mathfrak{H}$ having an orthonormal basis $\{e_i : i \geqslant 1\}$. For $q \geqslant 2$, denote by $\mathfrak{H}^{\otimes q}$ (resp. $\mathfrak{H}^{\odot q}$) the qth tensor power (resp. symmetric tensor power) of $\mathfrak{H}$; write moreover $\mathfrak{H}^{\otimes 0} = \mathfrak{H}^{\odot 0} = \mathbb{R}$ and $\mathfrak{H}^{\otimes 1} = \mathfrak{H}^{\odot 1} = \mathfrak{H}$. With every multi-index $\alpha \in \Lambda$, we associate the tensor $e(\alpha) \in \mathfrak{H}^{\otimes |\alpha|}$ given by

$$e(\alpha) = e_{i_1}^{\otimes \alpha_{i_1}} \otimes \cdots \otimes e_{i_k}^{\otimes \alpha_{i_k}},$$

where $\{\alpha_{i_1}, ..., \alpha_{i_k}\}$ are the non-zero elements of α. We also denote by $\tilde{e}(\alpha) \in \mathfrak{H}^{\odot|\alpha|}$ the canonical symmetrization of $e(\alpha)$. It is well-known that, for every $q \geqslant 2$, the set $\{\tilde{e}(\alpha) : \alpha \in \Lambda, |\alpha| = q\}$ is a complete orthogonal system in $\mathfrak{H}^{\odot q}$. For every $q \geqslant 1$ and every $h \in \mathfrak{H}^{\odot q}$ of the form $h = \sum_{\alpha\in\Lambda,\,|\alpha|=q} c_\alpha \tilde{e}(\alpha)$, we define

$$I_q(h) = \sum_{\alpha\in\Lambda,\,|\alpha|=q} c_\alpha \Phi(\alpha), \tag{3.60}$$

where $\Phi(\alpha)$ is given in (3.59). Another classical result (see e.g. [33, 38]) is that, for every $q \geqslant 1$, the mapping $I_q : \mathfrak{H}^{\odot q} \to C_q$ (as defined in (3.60)) is onto, and defines an isomorphism between C_q and the Hilbert space $\mathfrak{H}^{\odot q}$, endowed with the modified norm $\sqrt{q!}\|\cdot\|_{\mathfrak{H}^{\otimes q}}$. This means that, for every $h, h' \in \mathfrak{H}^{\odot q}$, $E[I_q(h)I_q(h')] = q!\langle h, h'\rangle_{\mathfrak{H}^{\otimes q}}$.

Finally, we observe that one can reexpress the Wiener-Itô chaotic decomposition of $L^2(\sigma(\mathbf{G}))$ as follows: every $F \in L^2(\sigma(\mathbf{G}))$ admits a unique decomposition of the type

$$F = \sum_{q=0}^{\infty} I_q(h_q), \tag{3.61}$$

where the series converges in $L^2(\mathbf{G})$, the symmetric kernels $h_q \in \mathfrak{H}^{\odot q}$, $q \geqslant 1$, are uniquely determined by F, and $I_0(h_0) := E[F]$. This also implies that

$$E[F^2] = E[F]^2 + \sum_{q=1}^{\infty} q!\|h_q\|^2_{\mathfrak{H}^{\otimes q}}.$$

3.2 The language of Malliavin calculus: chaoses as eigenspaces

We let the previous notation and assumptions prevail: in particular, we shall fix for the rest of the section a real separable Hilbert space $\mathfrak{H}$, and represent the elements of the qth Wiener chaos of $\mathbf{G}$ in the form (3.60). In addition to the previously introduced notation, $L^2(\mathfrak{H}) := L^2(\sigma(\mathbf{G}); \mathfrak{H})$ indicates the space of all $\mathfrak{H}$-valued random elements u, that are measurable with respect to $\sigma(\mathbf{G})$ and verify the relation $E\big[\|u\|^2_{\mathfrak{H}}\big] < \infty$. Note that, as it is customary, $\mathfrak{H}$ is endowed with the Borel σ-field associated with the distance on $\mathfrak{H}$ given by $(h_1, h_2) \mapsto \|h_1 - h_2\|_{\mathfrak{H}}$.

Let $\mathscr{S}$ be the set of all smooth cylindrical random variables of the form

$$F = g\big(I_1(h_1), \dots, I_1(h_n)\big),$$

where $n \geqslant 1$, $g : \mathbb{R}^n \to \mathbb{R}$ is a smooth function with compact support and $h_i \in \mathfrak{H}$. The Malliavin derivative of F (with respect to $\mathbf{G}$) is the element of $L^2(\mathfrak{H})$ defined as

$$DF \;=\; \sum_{i=1}^{n} \partial_i g\big(I_1(h_1), \dots, I_1(h_n)\big) h_i.$$

By iteration, one can define the mth derivative $D^m F$ (which is an element of $L^2(\mathfrak{H}^{\otimes m})$) for every $m \geqslant 2$. For $m \geqslant 1$, $\mathbb{D}^{m,2}$ denotes the closure of $\mathscr{S}$ with respect to the norm $\|\cdot\|_{m,2}$, defined by the relation

$$\|F\|^2_{m,2} = E[F^2] + \sum_{i=1}^{m} E\big[\|D^i F\|^2_{\mathfrak{H}^{\otimes i}}\big].$$

It is a standard result that a random variable F as in (3.61) is in $\mathbb{D}^{m,2}$ if and only if $\sum_{q\geqslant 1} q^m q!\|f_q\|^2_{\mathfrak{H}^{\otimes q}} < \infty$, from which one deduces that $\bigoplus_{k=0}^{q} C_k \in \mathbb{D}^{m,2}$ for every $q, m \geqslant 1$. Also, $DI_1(h) = h$ for every $h \in \mathfrak{H}$. The Malliavin derivative D satisfies the

following *chain rule*: if $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and has bounded partial derivatives, and if $(F_1, ..., F_n)$ is a vector of elements of $\mathbb{D}^{1,2}$, then $g(F_1, \ldots, F_n) \in \mathbb{D}^{1,2}$ and

$$Dg(F_1, \ldots, F_n) = \sum_{i=1}^{n} \partial_i g(F_1, \ldots, F_n) DF_i. \tag{3.62}$$

In what follows, we denote by δ the adjoint of the operator D, also called the *divergence operator*. A random element $u \in L^2(\mathfrak{H})$ belongs to the domain of δ, written $\operatorname{Dom}\delta$, if and only if it satisfies

$$|E[\langle DF, u\rangle_{\mathfrak{H}}]| \leqslant c_u \sqrt{E[F^2]} \quad \text{for any } F \in \mathscr{S},$$

for some constant c_u depending only on u. If $u \in \operatorname{Dom}\delta$, then the random variable $\delta(u)$ is defined by the duality relationship (customarily called "integration by parts formula"):

$$E[F\delta(u)] = E[\langle DF, u\rangle_{\mathfrak{H}}], \tag{3.63}$$

which holds for every $F \in \mathbb{D}^{1,2}$. A crucial object for our discussion is the Ornstein-Uhlenbeck semigroup associated with $\mathbf{G}$.

Definition 3.2 (Ornstein-Uhlenbeck semigroup) Let $\mathbf{G}'$ be an independent copy of $\mathbf{G}$, and denote by E' the mathematical expectation with respect to $\mathbf{G}'$. For every $t \geqslant 0$ the operator $P_t : L^2(\sigma(\mathbf{G})) \to L^2(\sigma(\mathbf{G}))$ is defined as follows: for every $F(\mathbf{G}) \in L^2(\sigma(\mathbf{G}))$,

$$P_t F(\mathbf{G}) = E'[F(e^{-t}\mathbf{G} + \sqrt{1 - e^{-2t}}\mathbf{G}')],$$

in such a way that $P_0 F(\mathbf{G}) = F(\mathbf{G})$ and $P_\infty F(\mathbf{G}) = E[F(\mathbf{G})]$. The collection $\{P_t : t \geqslant 0\}$ verifies the semigroup property $P_t P_s = P_{t+s}$ and is called the *Ornstein-Uhlenbeck semigroup* associated with $\mathbf{G}$.

The properties of the semigroup $\{P_t : t \geqslant 0\}$ that are relevant for our study are gathered together in the next statement.

Proposition 3.3

1. *For every $t > 0$, the eigenspaces of the operator P_t coincide with the Wiener chaoses C_q, $q = 0, 1, ...$, the eigenvalue of C_q being given by the positive constant e^{-qt}.*
2. *The* infinitesimal generator *of $\{P_t : t \geqslant 0\}$, denoted by L, acts on square-integrable random variables as follows: a random variable F with the form (3.61) is in the domain of L, written* $\operatorname{Dom} L$, *if and only if $\sum_{q\geqslant 1} qI_q(h_q)$ is convergent in $L^2(\sigma(\mathbf{G}))$, and in this case*

   $$LF = -\sum_{q\geqslant 1} qI_q(h_q).$$

   In particular, each Wiener chaos C_q is an eigenspace of L, with eigenvalue equal to $-q$.
3. *The operator L verifies the following properties:* (i) $\operatorname{Dom} L = \mathbb{D}^{2,2}$, *and* (ii) *a random variable F is in* $\operatorname{Dom} L$ *if and only if $F \in \operatorname{Dom}\delta D$ (i.e. $F \in \mathbb{D}^{1,2}$ and $DF \in \operatorname{Dom}\delta$), and in this case one has that $\delta(DF) = -LF$.*

In view of the previous statement, it is immediate to describe the *pseudo-inverse* of L, denoted by L^{-1}, as follows: for every mean zero random variable $F = \sum_{q\geqslant 1} I_q(h_q)$ of $L^2(\sigma(\mathbf{G}))$, one has that

$$L^{-1}F = \sum_{q\geqslant 1} -\frac{1}{q} I_q(h_q).$$

It is clear that L^{-1} is an operator with values in $\mathbb{D}^{2,2}$.

For future reference, we record the following estimate involving random variables living in a finite sum of Wiener chaoses: it is a direct consequence of the hypercontractivity of the Ornstein-Uhlenbeck semigroup – see e.g. in [33, Theorem 2.7.2 and Theorem 2.8.12].

Proposition 3.4 (Hypercontractivity) *Let $q \geqslant 1$ and $1 \leqslant s < t < \infty$. Then, there exists a finite constant $c(s,t,q) < \infty$ such that, for every $F \in \bigoplus_{k=0}^{q} C_k$,*

$$E[|F|^t]^{\frac{1}{t}} \leqslant c(s,t,q)\, E[|F|^s]^{\frac{1}{s}}. \tag{3.64}$$

In particular, all L^p norms, $p \geqslant 1$, are equivalent on a finite sum of Wiener chaoses.

Since we will systematically work on a fixed sum of Wiener chaoses, we will not need to specify the explicit value of the constant $c(s,t,q)$. See again [33], and the references therein, for more details.

Example 3.5 *Let $W = \{W_t : t \geqslant 0\}$ be a standard Brownian motion, let $\{e_j : j \geqslant 1\}$ be an orthonormal basis of $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), dt) =: L^2(\mathbb{R}_+)$, and define $G_j = \int_0^\infty e_j(t)dW_t$. Then, $\sigma(W) = \sigma(\mathbf{G})$, where $\mathbf{G} = \{G_j : j \geqslant 1\}$. In this case, the natural choice of a Hilbert space is $\mathfrak{H} = L^2(\mathbb{R}_+)$ and one has the following explicit characterisation of the Wiener chaoses associated with W: for every $q \geqslant 1$, one has that $F \in C_q$ if and only if there exists a symmetric kernel $f \in L^2(\mathbb{R}_+^q)$ such that*

$$F = q! \int_0^\infty \int_0^{t_1} \cdots \int_0^{t_{q-1}} f(t_1, ..., t_q) dW_{t_q} \cdots dW_{t_1} := q! J_q(f).$$

The random variable $J_q(f)$ is called the multiple Wiener-Itô integral *of order q, of f with respect to W. It is a well-known fact that, if $F \in \mathbb{D}^{1,2}$ admits the chaotic expansion $F = E[F] + \sum_{q \geqslant 1} J_q(f_q)$, then DF equals the random function*

$$\begin{aligned} t \mapsto & \sum_{q \geqslant 1} q J_{q-1}(f_q(t, \cdot)) \\ &= \sum_{q \geqslant 1} q! \int_0^\infty \int_0^{t_1} \cdots \int_0^{t_{q-2}} f(t, t_1 ..., t_{q-1}) dW_{t_{q-1}} \cdots dW_{t_1}, \quad t \in \mathbb{R}_+, \end{aligned}$$

which is a well-defined element of $L^2(\mathfrak{H})$.

3.3 The role of Malliavin and Stein matrices

Given a vector $F = (F_1, ..., F_d)$ whose elements are in $\mathbb{D}^{1,2}$, we define the *Malliavin matrix* $\Gamma(F) = \{\Gamma_{i,j}(F) : i, j = 1, ..., d\}$ as

$$\Gamma_{i,j}(F) = \langle DF_i, DF_j \rangle_{\mathfrak{H}}.$$

The following statement is taken from [31], and provides a simple necessary and sufficient condition for a random vector living in a finite sum of Wiener chaoses to have a density.

Theorem 3.6 *Fix $d, q \geqslant 1$ and let $F = (F_1, ..., F_d)$ be such that $F_i \in \bigoplus_{i=0}^{q} C_i$, $i = 1, ..., d$. Then, the distribution of F admits a density f with respect to the Lebesgue measure on $\mathbb{R}^d$ if and only if $E[\det \Gamma(F)] > 0$. Moreover, if this condition is verified one has necessarily that* $\mathrm{Ent}(F) < \infty$.

Proof. The equivalence between the existence of a density and the fact that $E[\det \Gamma(F)] > 0$ is shown in [31, Theorem 3.1]. Moreover, by [31, Theorem 4.2] we have in this case that the density f satisfies $f \in \bigcup_{p>1} L^p(\mathbb{R}^d)$. Relying on the inequality

$$\log u \leqslant n(u^{1/n} - 1), \quad u > 0, \quad n \in \mathbb{N}$$

(which is a direct consequence of $\log u \leqslant u - 1$ for any $u > 0$), one has that

$$\int_{\mathbb{R}^d} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} \leqslant n \left(\int_{\mathbb{R}^d} f(\mathbf{x})^{1+\frac{1}{n}} d\mathbf{x} - 1 \right).$$

Hence, by choosing n large enough so that $f \in L^{1+\frac{1}{n}}(\mathbb{R}^d)$, one obtains that $\mathrm{Ent}(F) < \infty$. ■

To conclude, we present a result providing an explicit representation for Stein matrices associated with random vectors in the domain of D.

Proposition 3.7 (Representation of Stein matrices) *Fix $d \geqslant 1$ and let $F = (F_1, ..., F_d)$ be a centered random vector whose elements are in $\mathbb{D}^{1,2}$. Then, a Stein matrix for F (see Definition 2.7) is given by*

$$\tau_F^{i,j}(\mathbf{x}) = E[\langle -DL^{-1}F_i, DF_j \rangle_{\mathfrak{H}} \big| F = \mathbf{x}], \quad i, j = 1, ..., d.$$

Proof. Let $g : \mathbb{R}^d \to \mathbb{R} \in C_c^1$. For every $i = 1, ..., d$, one has that $F_i = -\delta DL^{-1}F_i$. As a consequence, using (in order) (3.63) and (3.62), one infers that

$$E[F_i g(F)] = E[\langle -DL^{-1}F_i, Dg(F) \rangle_{\mathfrak{H}}] = \sum_{j=1}^{d} E[\langle -DL^{-1}F_i, DF_j \rangle_{\mathfrak{H}} \, \partial_j g(F)].$$

Taking conditional expectations yields the desired conclusion. ■

The next section contains the statements and proofs of our main bounds on a Gaussian space.

4 Entropic fourth moment bounds on a Gaussian space

4.1 Main results

We let the framework of the previous section prevail: in particular, $\mathbf{G}$ is a sequence of i.i.d. $\mathscr{N}(0,1)$ random variables, and the sequence of Wiener chaoses $\{C_q : q \geqslant 1\}$ associated with $\mathbf{G}$ is encoded by means of increasing tensor powers of a fixed real separable Hilbert space $\mathfrak{H}$. We will use the following notation: given a sequence of centered and square-integrable d-dimensional random vectors $F_n = (F_{1,n}, ..., F_{d,n}) \in \mathbb{D}^{1,2}$ with covariance matrix C_n, $n \geqslant 1$, we let

$$\Delta_n := E[\|F_n\|^4] - E[\|Z_n\|^4], \tag{4.65}$$

where $\| \cdot \|$ is the Euclidian norm on $\mathbb{R}^d$ and $Z_n \sim \mathscr{N}_d(0, C_n)$.

Our main result is the following entropic central limit theorem for sequences of chaotic random variables.

Theorem 4.1 (Entropic CLTs on Wiener chaos) *Let $d \geqslant 1$ and $q_1, \ldots, q_d \geqslant 1$ be fixed integers. Consider vectors*

$$F_n = (F_{1,n}, \ldots, F_{d,n}) = (I_{q_1}(h_{1,n}), \ldots, I_{q_d}(h_{d,n})), \quad n \geqslant 1,$$

with $h_{i,n} \in \mathfrak{H}^{\odot q_i}$. Let C_n denote the covariance matrix of F_n and let $Z_n \sim \mathscr{N}_d(0, C_n)$ be a centered Gaussian random vector in $\mathbb{R}^d$ with the same covariance matrix as F_n. Assume that $C_n \to C > 0$ and $\Delta_n \to 0$, as $n \to \infty$. Then, the random vector F_n admits a density for n large enough, and

$$D(F_n \| Z_n) = O(1)\, \Delta_n |\log \Delta_n| \quad \text{as } n \to \infty, \tag{4.66}$$

where $O(1)$ indicates a bounded numerical sequence depending on $d, q_1, ..., q_d$, as well as on the sequence $\{F_n\}$.

One immediate consequence of the previous statement is the following characterisation of entropic CLTs on a finite sum of Wiener chaoses.

Corollary 4.2 *Let the sequence $\{F_n\}$ be as in the statement of Theorem 4.1, and assume that $C_n \to C > 0$. Then, the following three assertions are equivalent, as $n \to \infty$:*

(i) $\Delta_n \to 0$;

(ii) *F_n converges in distribution to $Z \sim \mathscr{N}_d(0, C)$;*

(iii) $D(F_n \| Z_n) \to 0$.

The proofs of the previous results are based on two technical lemmas that are the object of the next section.

4.2 Estimates based on the Carbery-Wright inequalities

We start with a generalisation of the inequality proved by Carbery and Wright in [11]. Recall that, in a form that is adapted to our framework, the main finding of [11] reads as follows: there is a universal constant $c > 0$ such that, for any polynomial $Q : \mathbb{R}^n \to \mathbb{R}$ of degree at most d and any $\alpha > 0$ we have

$$E[Q(X_1, \ldots, X_n)^2]^{\frac{1}{2d}}\ P(|Q(X_1, \ldots, X_n)| \leqslant \alpha) \leqslant c d \alpha^{\frac{1}{d}}, \tag{4.67}$$

where $X_1, \ldots, X_n$ are independent random variables with common distribution $\mathscr{N}(0, 1)$.

Lemma 4.3 *Fix $d, q_1, \ldots, q_d \geqslant 1$, and let $F = (F_1, \ldots, F_d)$ be a random vector such that $F_i = I_{q_i}(h_i)$ with $h_i \in \mathfrak{H}^{\odot q_i}$. Let $\Gamma = \Gamma(F)$ denote the Malliavin matrix of F, and assume that $E[\det \Gamma] > 0$ (which is equivalent to assuming that F has a density by Theorem 3.6). Set $N = 2d(q-1)$ with $q = \max_{1 \leqslant i \leqslant d} q_i$. Then, there exists a universal constant $c > 0$ such that*

$$P(\det \Gamma \leqslant \lambda) \leqslant c N \lambda^{1/N} (E[\det \Gamma])^{-1/N}. \tag{4.68}$$

Proof. Let $\{e_i : i \geqslant 1\}$ be an orthonormal basis of $\mathfrak{H}$. Since $\det \Gamma$ is a polynomial of degree d in the entries of Γ and because each entry of Γ belongs to $\bigoplus_{k=0}^{2q-2} C_k$ by the product formula for multiple integrals (see, e.g., [33, Chapter 2]), we have, by iterating the product formula, that $\det \Gamma \in \bigoplus_{k=0}^{N} C_k$. Thus, there exists a sequence $\{Q_n, n \geqslant 1\}$ of real-valued polynomials of degree at most N such that the random variables $Q_n(I_1(e_1), \ldots, I_1(e_n))$ converge in L^2 and almost surely to $\det \Gamma$ as n tends to infinity (see [36, proof of Theorem 3.1] for an explicit construction). Assume now that $E[\det \Gamma] > 0$. Then, for n sufficiently large, $E[|Q_n(I_1(e_1), \ldots, I_1(e_n))|] > 0$. We deduce from the estimate (4.67) the existence of a universal constant $c > 0$ such that, for any $n \geqslant 1$,

$$P(|Q_n(I_1(e_1), \ldots, I_1(e_n)| \leqslant \lambda) \leqslant c N \lambda^{1/N} (E[Q_n(I_1(e_1), \ldots, I_1(e_n)^2])^{-1/2N}.$$

Using the property

$$E[Q_n(I_1(e_1), \ldots, I_1(e_n)^2] \geqslant (E[|Q_n(I_1(e_1), \ldots, I_1(e_n)|])^2$$

we obtain

$$P(|Q_n(I_1(e_1),\dots,I_1(e_n)| \leqslant \lambda) \leqslant cN\lambda^{1/N}(E[|Q_n(I_1(e_1),\dots,I_1(e_n))|])^{-1/N},$$

from which (4.68) follows by letting n tend to infinity. ■

The next statement, whose proof follows the same lines than that of [31, Theorem 4.1], provides an upper bound on the total variation distance between the distribution of F and that of $\sqrt{t}F+\sqrt{1-t}\mathbf{x}$, for every $\mathbf{x}\in\mathbb{R}^d$ and every $t\in[1/2;1]$. Although Lemma 4.4 is, in principle, very close to [31, Theorem 4.1], we detail its proof because, here, we need to keep track of the way the constants behave with respect to $\mathbf{x}$.

Lemma 4.4 *Fix $d, q_1,\dots,q_d \geqslant 1$, and let $F=(F_1,\dots,F_d)$ be a random vector as in Lemma 4.3. Set $q=\max_{1\leqslant i\leqslant d} q_i$. Let C be the covariance matrix of F, let $\Gamma=\Gamma(F)$ denote the Malliavin matrix of F, and assume that $\beta := E[\det\Gamma]>0$. Then, there exists a constant $c_{q,d,\|C\|_{H.S.}}>0$ (depending only on q, d and $\|C\|_{H.S.}$ — with a continuous dependence in the last parameter) such that, for any $\mathbf{x}\in\mathbb{R}^d$ and $t\in[\frac{1}{2},1]$,*

$$\mathbf{TV}(\sqrt{t}F+\sqrt{1-t}\mathbf{x},F) \leqslant c_{q,d,\|C\|_{H.S.}}\left(\beta^{-\frac{1}{N+1}}\wedge 1\right)(1+\|\mathbf{x}\|_1)\,(1-t)^{\frac{1}{2(2N+4)(d+1)+2}}. \quad (4.69)$$

Here, $N=2d(q-1)$.

Proof. The proof is divided into several steps. In what follows, we fix $t\in[\frac{1}{2},1]$ and $\mathbf{x}\in\mathbb{R}^d$ and we use the convention that $c_{\{.\}}$ denotes a constant in $(0,\infty)$ that only depends on the arguments inside the bracket and whose value is allowed to change from one line to another.

Step 1. One has that $E\big[\|F\|_\infty^2\big] \leqslant c_d E\big[\|F\|_2^2\big] \leqslant c_d\sqrt{\sum_{i,j=1}^d C(i,i)^2} \leqslant c_d\|C\|_{H.S.}$ so that $E\big[\|F\|_\infty\big]\leqslant c_d\sqrt{\|C\|_{H.S.}}=c_{d,\|C\|_{H.S.}}$ Let $g:\mathbb{R}^d\to\mathbb{R}$ be a (smooth) test function bounded by 1. We can write, for any $M\geqslant 1$,

$$\begin{aligned}
&\big|E[g(\sqrt{t}F+\sqrt{1-t}\mathbf{x})]-E[g(F)]\big| \\
&\leqslant \Big|E\left[(g\mathbf{1}_{[-M/2,M/2]^d})(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right]-E\left[(g\mathbf{1}_{[-M/2,M/2]^d})(F)\right]\Big| \\
&\quad + P(\|\sqrt{t}F+\sqrt{1-t}\mathbf{x}\|_\infty\geqslant M/2)+P(\|F\|_\infty\geqslant M/2) \\
&\leqslant \sup_{\substack{\|\phi\|_\infty\leqslant 1\\ \mathrm{supp}\phi\subset[-M,M]^d}} \big|E[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})]-E[\phi(F)]\big| \\
&\quad + \frac{2}{M}\left(E\left[\|\sqrt{t}F+\sqrt{1-t}\mathbf{x}\|_\infty\right]+E\left[\|F\|_\infty\right]\right) \\
&\leqslant \sup_{\substack{\|\phi\|_\infty\leqslant 1\\ \mathrm{supp}\phi\subset[-M,M]^d}} \big|E[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})]-E[\phi(F)]\big| + \frac{c_{d,\|C\|_{H.S.}}}{M}(1+\|\mathbf{x}\|_\infty). \quad (4.70)
\end{aligned}$$

Step 2. Let $\phi:\mathbb{R}^d\to\mathbb{R}$ be $\mathcal{C}^\infty$ with compact support in $[-M,M]^d$ and satisfying $\|\phi\|_\infty\leqslant 1$. Let $0<\alpha\leqslant 1$ and let $\rho:\mathbb{R}^d\to\mathbb{R}_+$ be in $\mathcal{C}_c^\infty$ and satisfying $\int_{\mathbb{R}^d}\rho(x)dx=1$. Set $\rho_\alpha(x)=\frac{1}{\alpha^d}\rho(\frac{x}{\alpha})$. By [36, formula (3.26)], we have that $\phi*\rho_\alpha$ is Lipschitz continuous

with constant $1/\alpha$. We can thus write,

$$
\begin{aligned}
&\left|E[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})]-E[\phi(F)]\right| \\
&\leqslant \left|E\left[\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(F)\right]\right|+\left|E\left[\phi(F)-\phi*\rho_\alpha(F)\right]\right| \\
&\quad+\left|E\left[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right]\right| \\
&\leqslant \frac{\sqrt{1-t}}{\alpha}\left(E\left[\|F\|_\infty\right]+\|\mathbf{x}\|_\infty\right)+\left|E\left[\phi(F)-\phi*\rho_\alpha(F)\right]\right| \\
&\quad+\left|E\left[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right]\right| \\
&\leqslant c_{d,\|C\|_{H.S.}}\frac{\sqrt{1-t}}{\alpha}\left(1+\|\mathbf{x}\|_1\right)+\left|E\left[\phi(F)-\phi*\rho_\alpha(F)\right]\right| \\
&\quad+\left|E\left[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right]\right|. \qquad (4.71)
\end{aligned}
$$

In order to estimate the two last terms in (4.71), we decompose the expectation into two parts using the identity

$$
1=\frac{\varepsilon}{t^d\det\Gamma+\varepsilon}+\frac{t^d\det\Gamma}{t^d\det\Gamma+\varepsilon},\quad \varepsilon>0.
$$

Step 3. For all $\varepsilon,\lambda>0$ and using (4.68),

$$
\begin{aligned}
&E\left[\frac{\varepsilon}{t^d\det\Gamma+\varepsilon}\right] \\
&\leqslant E\left[\frac{\varepsilon}{t^d\det\Gamma+\varepsilon}\mathbf{1}_{\{\det\Gamma>\lambda\}}\right]+cN\left(\frac{\lambda}{E[\det\Gamma]}\right)^{1/N} \\
&\leqslant \frac{\varepsilon}{t^d\lambda}+cN\left(\frac{\lambda}{\beta}\right)^{1/N}.
\end{aligned}
$$

Choosing $\lambda=\varepsilon^{\frac{N}{N+1}}\beta^{\frac{1}{N+1}}$ yields

$$
E\left[\frac{\varepsilon}{t^d\det\Gamma+\varepsilon}\right]\leqslant(t^{-d}+cN)\left(\frac{\varepsilon}{\beta}\right)^{\frac{1}{N+1}}\leqslant c_{q,d}\left(\frac{\varepsilon}{\beta}\right)^{\frac{1}{N+1}}\qquad\text{(recall that } t\geqslant\tfrac{1}{2}\text{)}.
$$

As a consequence,

$$
\begin{aligned}
&\left|E\left[\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right]\right| \\
&=\left|E\left[\left(\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right)\times\right.\right. \\
&\quad\left.\left.\times\left(\frac{\varepsilon}{t^d\det\Gamma+\varepsilon}+\frac{t^d\det\Gamma}{t^d\det\Gamma+\varepsilon}\right)\right]\right| \\
&\leqslant c_{q,d}\left(\frac{\varepsilon}{\beta}\right)^{\frac{1}{N+1}} \\
&\quad+\left|E\left[\left(\phi(\sqrt{t}F+\sqrt{1-t}\mathbf{x})-\phi*\rho_\alpha(\sqrt{t}F+\sqrt{1-t}\mathbf{x})\right)\frac{t^d\det\Gamma}{t^d\det\Gamma+\varepsilon}\right]\right|. \qquad (4.72)
\end{aligned}
$$

Step 4. In this step we recall from [31, page 11] the following integration by parts formula (4.73). Let $h:\mathbb{R}^d\to\mathbb{R}$ be a function in $\mathcal{C}^\infty$ with compact support, and consider a

random variable $W \in \mathbb{D}^\infty$. Consider the Poisson kernel in $\mathbb{R}^d$, defined as the solution to the equation $\Delta Q_d = \delta_0$. We know that $Q_2(x) = c_2 \log|x|$ and $Q_d(x) = c_d |x|^{2-d}$ for $d \neq 2$. We have the following identity

$$
\begin{aligned}
&E[W \det\Gamma\, h(\sqrt{t}F + \sqrt{1-t}\mathbf{x})] \\
&= \frac{1}{\sqrt{t}} \sum_{i=1}^{d} E\left[A_i(W) \int_{\mathbb{R}^d} h(y) \partial_i Q_d(\sqrt{t}F + \sqrt{1-t}\mathbf{x} - y) dy \right],
\end{aligned}
\tag{4.73}
$$

where

$$
A_i(W) = -\sum_{a=1}^{d} \big(\langle D(W(\mathrm{Adj}\Gamma)_{a,i}), DF_a \rangle_{\mathfrak{H}} + (\mathrm{Adj}\Gamma)_{a,i} W LF_a \big),
$$

with Adj the usual adjugate matrix operator.

Step 5. Let us apply the identity (4.73) to the function $h = \phi - \phi * \rho_\alpha$ and to the random variable $W = W_\varepsilon = \frac{1}{t^d \det\Gamma + \varepsilon} \in \mathbb{D}^\infty$; we obtain

$$
\begin{aligned}
&E\left[(\phi - \phi * \rho_\alpha)(\sqrt{t}F + \sqrt{1-t}\mathbf{x}) \frac{t^d \det\Gamma}{t^d \det\Gamma + \varepsilon} \right] \\
&= t^{d-\frac{1}{2}} \sum_{i=1}^{d} E\left[A_i(W_\varepsilon) \int_{\mathbb{R}^d} (\phi - \phi * \rho_\alpha)(\mathbf{y}) \partial_i Q_d(\sqrt{t}F + \sqrt{1-t}\mathbf{x} - \mathbf{y}) d\mathbf{y} \right].
\end{aligned}
\tag{4.74}
$$

From the hypercontractivity property together with the equality

$$
\begin{aligned}
A_i(W_\varepsilon) = \sum_{a=1}^{d} \Bigg\{ &- \frac{1}{t^d \det\Gamma + \varepsilon} \left(\langle D(\mathrm{Adj}\Gamma)_{a,i}, DF_a \rangle_{\mathfrak{H}} - (\mathrm{Adj}\Gamma)_{a,i} LF_a \right) \\
&+ \frac{1}{(t^d \det\Gamma + \varepsilon)^2} (\mathrm{Adj}\Gamma)_{a,i} \langle D(\det\Gamma), DF_a \rangle_{\mathfrak{H}} \Bigg\},
\end{aligned}
$$

one immediately deduces the existence of $c_{q,d,\|C\|_{H.S.}} > 0$ such that

$$
E[A_i(W_\varepsilon)^2] \leqslant c_{q,d,\|C\|_{H.S.}}\, \varepsilon^{-4}.
$$

On the other hand, in [31, page 13] it is shown that there exists $c_d > 0$ such that, for any $R > 0$ and $u \in \mathbb{R}^d$,

$$
\left| \int_{\mathbb{R}^d} (\phi - \phi * \rho_\alpha)(\mathbf{y}) \partial_i Q_d(\mathbf{u} - \mathbf{y}) d\mathbf{y} \right| \leqslant c_d \left(R + \alpha + \alpha R^{-d} (\|\mathbf{u}\|_1 + M)^d \right).
$$

Substituting this estimate into (4.74) and assuming that $M \geqslant 1$, yields

$$
\begin{aligned}
&\left| E\left[(\phi - \phi * \rho_\alpha)(\sqrt{t}F + \sqrt{1-t}\mathbf{x}) \frac{t^d \det\Gamma}{t^d \det\Gamma + \varepsilon} \right] \right| \\
&\leqslant c_{q,d,\|C\|_{H.S.}}\, \varepsilon^{-2} \left(R + \alpha + \alpha R^{-d} (M + \|\mathbf{x}\|_1)^d \right).
\end{aligned}
$$

Choosing $R = \alpha^{\frac{1}{d+1}} (M + \|\mathbf{x}\|_1)^{\frac{d}{d+1}}$ and assuming $\alpha \leqslant 1$, we obtain

$$
\begin{aligned}
&\left| E\left[(\phi - \phi * \rho_\alpha)(\sqrt{t}F + \sqrt{1-t}\mathbf{x}) \frac{t^d \det\Gamma}{t^d \det\Gamma + \varepsilon} \right] \right| \\
&\leqslant c_{q,d,\|C\|_{H.S.}}\, \varepsilon^{-2} \alpha^{\frac{1}{d+1}} (M + \|\mathbf{x}\|_1)^{\frac{d}{d+1}}.
\end{aligned}
\tag{4.75}
$$

It is worthwhile noting that the inequality (4.75) is valid for any $t \in [\frac{1}{2}, 1]$, in particular for $t = 1$.

Step 6. From (4.71), (4.72) and (4.75) we obtain

$$
\begin{aligned}
&|E[\phi(\sqrt{t}F + \sqrt{1-t}\mathbf{x})] - E[\phi(F)]| \\
&\leqslant c_{d,\|C\|_{H.S.}} \frac{\sqrt{1-t}}{\alpha}(1 + \|\mathbf{x}\|_1) + c_{q,d}\left(\frac{\varepsilon}{\beta}\right)^{\frac{1}{N+1}} \\
&\quad + c_{q,d,\|C\|_{H.S.}}\, \varepsilon^{-2}\alpha^{\frac{1}{d+1}}(M + \|\mathbf{x}\|_1)^{\frac{d}{d+1}}.
\end{aligned}
$$

By plugging this inequality into (4.70) we thus obtain that, for every $M \geqslant 1$, $\varepsilon > 0$ and $0 < \alpha \leqslant 1$:

$$
\begin{aligned}
&\mathbf{TV}(\sqrt{t}F + \sqrt{1-t}\mathbf{x}, F) \\
&\leqslant c_{d,\|C\|_{H.S.}} \frac{\sqrt{1-t}}{\alpha}(1 + \|\mathbf{x}\|_1) + c_{q,d}\left(\frac{\varepsilon}{\beta}\right)^{\frac{1}{N+1}} + c_{q,d,\|C\|_{H.S.}} \varepsilon^{-2}\alpha^{\frac{1}{d+1}}(M + \|\mathbf{x}\|_1)^{\frac{d}{d+1}} \\
&\quad + \frac{c_{d,\|C\|_{H.S.}}}{M}(1 + \|\mathbf{x}\|_1) \\
&\leqslant c_{q,d,\|C\|_{H.S.}}(1 + \|\mathbf{x}\|_1)\left(\beta^{-\frac{1}{N+1}} \wedge 1\right)\left\{\frac{\sqrt{1-t}}{\alpha} + \varepsilon^{\frac{1}{N+1}} + \frac{\alpha^{\frac{1}{d+1}}M}{\varepsilon^2} + \frac{1}{M}\right\}
\end{aligned}
\tag{4.76}
$$

Choosing $M = \varepsilon^{-\frac{1}{N+1}}$, $\varepsilon = \alpha^{\frac{N+1}{(2N+4)(d+1)}}$ and $\alpha = (1-t)^{\frac{(2N+4)(d+1)}{2((2N+4)(d+1)+1)}}$, one obtains the desired conclusion (4.69). ■

4.3 Proof of Theorem 4.1

In the proof of [37, Theorem 4.3], the following two facts have been shown:

$$
E\left[\left(C_n(j,k) - \frac{1}{q_j}\langle DF_{j,n}, DF_{k,n}\rangle_{\mathfrak{H}}\right)^2\right] \leqslant \mathrm{Cov}(F_{j,n}^2, F_{k,n}^2) - 2C_n(j,k)^2
$$

$$
E\left[\|F_n\|^4\right] - E\left[\|Z_n\|^4\right] = \sum_{j,k=1}^d \left\{\mathrm{Cov}(F_{j,n}^2, F_{k,n}^2) - 2C_n(j,k)^2\right\}.
$$

As a consequence, one deduces that

$$
\sum_{j,k=1}^d E\left[\left(C_n(j,k) - \frac{1}{q_j}\langle DF_{j,n}, DF_{k,n}\rangle_{\mathfrak{H}}\right)^2\right] \leqslant \Delta_n.
$$

Using Proposition 3.7, one infers immediately that

$$
\tau_{F_n}^{j,k}(\mathbf{x}) := \frac{1}{q_j}E[\langle DF_{j,n}, DF_{k,n}\rangle_{\mathfrak{H}} | F_n = \mathbf{x}], \quad j,k = 1,\ldots,d, \tag{4.77}
$$

defines a Stein's matrix for F_n, which moreover satisfies the relation

$$
\sum_{j,k=1}^d E\left[\left(C(j,k) - \tau_{F_n}^{j,k}(F_n)\right)^2\right] \leqslant \Delta_n. \tag{4.78}
$$

Now let Γ_n denote the Malliavin matrix of F_n. Thanks to [39, Lemma 6], we know that, for any $i,j = 1,\ldots,d$,

$$
\langle DF_{i,n}, DF_{j,n}\rangle_{\mathfrak{H}} \overset{L^2(\sigma(\mathbf{G}))}{\longrightarrow} \sqrt{q_i q_j}C(i,j) \quad \text{as } n \to \infty. \tag{4.79}
$$

Since $\langle DF_{i,n}, DF_{j,n}\rangle_{\mathfrak{H}}$ lives in a finite sum of chaoses (see e.g. [33, Chapter 5]) and is bounded in $L^2(\sigma(\mathbf{G}))$, we can again apply the hypercontractive estimate (3.64) to deduce that $\langle DF_{i,n}, DF_{j,n}\rangle_{\mathfrak{H}}$ is actually bounded in $L^p(\sigma(\mathbf{G}))$ for every $p \geqslant 1$, so that the convergence in (4.79) is in the sense of any of the spaces $L^p(\sigma(\mathbf{G}))$. As a consequence, $E[\det \Gamma_n] \to \det C \prod_{i=1}^d q_i =: \gamma > 0$, and there exists n_0 large enough so that

$$\inf_{n\geqslant n_0} E[\det \Gamma_n] > 0.$$

We are now able to deduce from Lemma 4.4 the existence of two constants $\kappa > 0$ and $\alpha \in (0, \frac{1}{2}]$ such that, for all $\mathbf{x} \in \mathbb{R}^d$, $t \in [\frac{1}{2}, 1]$ and $n \geqslant n_0$,

$$\mathbf{TV}(\sqrt{t}F_n + \sqrt{1-t}\mathbf{x}, F_n) \leqslant \kappa(1 + \|\mathbf{x}\|_1)(1-t)^\alpha.$$

This means that relation (2.53) is satisfied uniformly on n. Concerning (2.52), again by hypercontractivity and using the representation (4.77), one has that, for all $\eta > 0$,

$$\sup_{n\geqslant 1} E\big[|\tau_{F_n}^{j,k}(F_n)|^{\eta+2}\big] < \infty, \quad j, k = 1, \ldots, d.$$

Finally, since $\Delta_n \to 0$ and because (4.78) holds true, the condition (2.54) is satisfied for n large enough. The proof of (4.66) is concluded by applying Theorem 2.11.

4.4 Proof of Corollary 4.2

In view of Theorem 4.1, one has only to prove that (b) implies (a). This is an immediate consequence of the fact that the covariance C_n converges to C, and that the sequence $\{F_n\}$ lives in a finite sum of Wiener chaoses. Indeed, by virtue of (3.64) one has that

$$\sup_{n\geqslant 1} E\left[\|F_n\|^p\right] < \infty, \quad \forall p \geqslant 1,$$

yielding in particular that, if F_n converges in distribution to Z, then $E\|F_n\|^4 \to E\|Z\|^4$ or, equivalently, $\Delta_n \to 0$. The proof is concluded.

Acknowledgement. We heartily thank Guillaume Poly for several crucial inputs that helped us achieving the proof of Theorem 4.1. We are grateful to Larry Goldstein and Oliver Johnson for useful discussions. Ivan Nourdin is partially supported by the ANR grant ANR-10-BLAN-0121.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[2] S. Artstein, K. Ball, F. Barthe, and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. *Journal of the American Mathematical Society*, 17(4):975–982, 2004.

[3] S. Artstein, K. M. Ball, F. Barthe, and A. Naor. On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields*, 129(3):381–390, 2004.

[4] D. Bakry and M. Émery. Diffusions hypercontractives. In *Lecture Notes in Math.*, number 1123 in Séminaire de Probabilités XIX, pages 179–206. Springer, 1985.

[5] K. Ball, F. Barthe, and A. Naor. Entropy jumps in the presence of a spectral gap. *Duke Math. J.*, 119(1):41–63, 2003.

[6] K. Ball and V. Nguyen. Entropy jumps for random vectors with log-concave density and spectral gap. Preprint, arxiv:1206.5098v3, 2012.

[7] A. D. Barbour, O. Johnson, I. Kontoyiannis, and M. Madiman. Compound Poisson approximation via information functionals. *Electron. J. Probab.*, 15:1344–1368, 2010.

[8] A. R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14(1):336–342, 1986.

[9] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Fisher information and the central limit theorem. Preprint, arXiv:1204.6650v1, 2012.

[10] L. D. Brown. A proof of the central limit theorem motivated by the Cramér-Rao inequality. In *Statistics and probability: essays in honor of C. R. Rao*, pages 141–148. North-Holland, Amsterdam, 1982.

[11] A. Carbery and J. Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Res. Lett.*, 8(3):233–248, 2001.

[12] E. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorems. *Commun. Math. Phys.*, 140(2):339–371, 1991.

[13] S. Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40, 2009.

[14] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008.

[15] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[16] I. Csiszár. Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 7:137–158, 1962.

[17] A. Deya, S. Noreddine, and I. Nourdin. Fourth moment theorem and q-Brownian chaos. *Comm. Math. Phys., to appear*, 02 2012.

[18] R. M. Dudley. *Real analysis and probability*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[19] O. Johnson. *Information theory and the central limit theorem*. Imperial College Press, London, 2004.

[20] O. Johnson and A. Barron. Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, 129(3):391–409, 2004.

[21] O. Johnson and Y. Suhov. Entropy and random vectors. *Journal of Statistical Physics*, 104(1):145–165, 2001.

[22] T. Kemp, I. Nourdin, G. Peccati, and R. Speicher. Wigner chaos and the fourth moment. *Ann. Probab.*, 40(4):1577–1635, 2012.

[23] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Info. Theory*, 4, 1967.

[24] S. Kumar Kattumannil. On Stein's identity and its applications. *Statist. Probab. Lett.*, 79(12):1444–1449, 2009.

[25] M. Ledoux. Chaos of a Markov operator and the fourth moment condition. *Ann. Probab.*, 40(6):2439–2459, 2012.

[26] C. Ley and Y. Swan. Stein's density approach for discrete distributions with applications to information inequalities. Preprint, arXiv:1211.3668, 2012.

[27] C. Ley and Y. Swan. Stein's density approach and information inequalities. *Electron. Comm. Probab.*, 18(7):1–14, 2013.

[28] J. V. Linnik. An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Theor. Probability Appl.*, 4:288–299, 1959.

[29] D. Marinucci and G. Peccati. *Random fields on the sphere : Representation, limit theorems and cosmological applications*, volume 389 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2011.

[30] I. Nourdin. *Selected Aspects of Fractional Brownian Motion*. Springer-Verlag, 2012.

[31] I. Nourdin, D. Nualart, and G. Poly. Absolute continuity and convergence of densities for random vectors on Wiener chaos. *Electron. J. Probab.*, 18(22):1–19, 2012.

[32] I. Nourdin and G. Peccati. Universal Gaussian fluctuations of non-Hermitian matrix ensembles: from weak convergence to almost sure CLTs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 7:341–375, 2010.

[33] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus : from Stein's method to universality*. Cambridge Tracts in Mathematics. Cambridge University Press, 2012.

[34] I. Nourdin, G. Peccati, and G. Reinert. Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *Ann. Probab.*, 38(5):1947–1985, 2010.

[35] I. Nourdin, G. Peccati, and A. Réveillac. Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):45–58, 2010.

[36] I. Nourdin and G. Poly. Convergence in total variation on Wiener chaos. *Stochastic Process. Appl.*, to appear, 2012.

[37] I. Nourdin and J. Rosiński. Asymptotic independence of multiple Wiener-Itô integrals and the resulting limit laws. *Ann. Probab.*, to appear, 2013.

[38] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[39] D. Nualart and S. Ortiz-Latorre. Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stochastic Process. Appl.*, 118(4):614–628, 2008.

[40] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005.

[41] S. Park, E. Serpedin, and K. Qaraqe. On the equivalence between Stein and de Bruijn identities. *IEEE Trans. Info. Theory*, 58(12):7045–7067, 2012.

[42] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated and edited by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif., 1964.

[43] G. Reinert and A. Roellin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.*, 37(6):2150–2173, 2009.

[44] I. Sason. Entropy bounds for discrete random variables via coupling. Preprint, arXiv:1209.5259, 2012.

[45] I. Sason. An information-theoretic perspective of the Poisson approximation via the Chen-Stein method. Preprint, arXiv:1206.6811v3, 2012.

[46] R. Shimizu. On Fisher's amount of information for location family. In *A Modern Course on Statistical Distributions in Scientific Work*, pages 305–312. Springer, 1975.

[47] C. Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

[48] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

[49] F. G. Viens. Stein's lemma, Malliavin calculus, and tail bounds, with application to polymer fluctuation exponent. *Stochastic Process. Appl.*, 119(10):3671–3698, 2009.